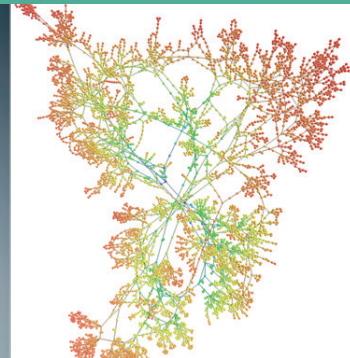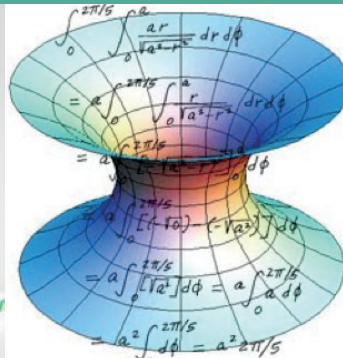# 5TH URV DOCTORAL WORKSHOP IN COMPUTER SCIENCE AND MATHEMATICS

Edited by Antoni Martínez–Ballesté, Agustí Solanas & Aïda Valls Mateu

UNIVERSITAT ROVIRA i VIRGILI

Title: 5rthURV Doctoral Workshop in Computer Science and Mathematics
Editors: Antoni Martínez-Ballesté, Agustí Solanas & Aïda Valls Mateu
May 2019

Preface

This book of proceedings gathers the contributions presented at the *5th URV Doctoral Workshop in Computer Science and Mathematics*. This edition has been held in Tarragona (Catalonia, Spain) on May 22nd, 2019. It has been jointly organized by the research group Smart Health and the Doctoral Program on Computer Science and Mathematics of Security of Universitat Rovira i Virgili (URV). The main aim of this workshop is to promote the dissemination of the ideas, methods and results that are developed in the Doctoral Thesis of the students of this doctorate program, and to promote the knowledge, collaboration and discussion between their respective research groups.

In this book, the reader will find the contributions of the Ph.D. students. Each chapter presents the research topic of one student, the goals and some of the results. The editors and organizers invite you to contact the authors for more detailed explanations and we encourage you to send them your suggestions and comments that may certainly help them in the next steps of their PhD thesis.

We thank all the participants and, especially, the students that presented their work in this DCSM workshop. Finally, we also want to thank Universitat Rovira i Virgili (URV), the Departament d'Enginyeria Informàtica i Matemàtiques (DEIM), and the Escola Tècnica Superior d'Enginyeria (ETSE) for their support.

Antoni Martínez-Ballesté, Agusti Solanas and Aïda Valls

Contents

Part I

Oral Presentations

Fast Text Localization in Images Using a Linear Spatial Filter

Xavier Gironés *

Department of Computer Science and Mathematics
Rovira i Virgili University, Tarragona, Spain
`xavier.girones@urv.cat`

Abstract. This paper proposes a novel text localization method in natural images based on the connected components (CC) approach. First, CC are isolated by convolving a multi-scale pyramid with a specifically designed linear spatial filter followed by hysteresis thresholding. Next, non-textual CC are pruned employing a local classifier consisting of a cascade of multilayer perceptron (MLP) fed with increasingly extended feature vectors. The stroke width feature is estimated in linear time complexity by computing the maximal inscribed squares in the CC. Candidate CC and their neighbors are then checked using a more global MLP classifier that takes into account the target CC and their vicinity. Finally, text sequences are extracted in all pyramid levels and fused using dynamic programming. The main contribution of the proposed method is its execution speed, being capable of processing 1080p HD video at nearly 30 frames per second on a standard laptop. In addition, it delivers competitive results in terms of precision and recall on the ICDAR 2013 Robust Reading dataset.

1 Introduction

While the topic of *optical character recognition* (OCR) on scanned documents has been intensively studied during the past decades and has attained a degree of maturity, the problem of text detection and recognition in natural scene images, *text spotting*, still remains a challenge. Factors present in natural images such as background clutter, occlusions, poor lighting conditions, shadows, perspective distortions, blurring, variation in font, scale, and orientation make the task of text spotting considerably more difficult that the typical OCR operation. There are two main categories of text detection methods described in the literature. On the one hand, texture based methods scan the image using a sliding window, typically at different scales. Texture properties are extracted in a feature vector and then fed to a classifier. As a final step, neighboring text regions are merged and combined into text lines. The advantages of these
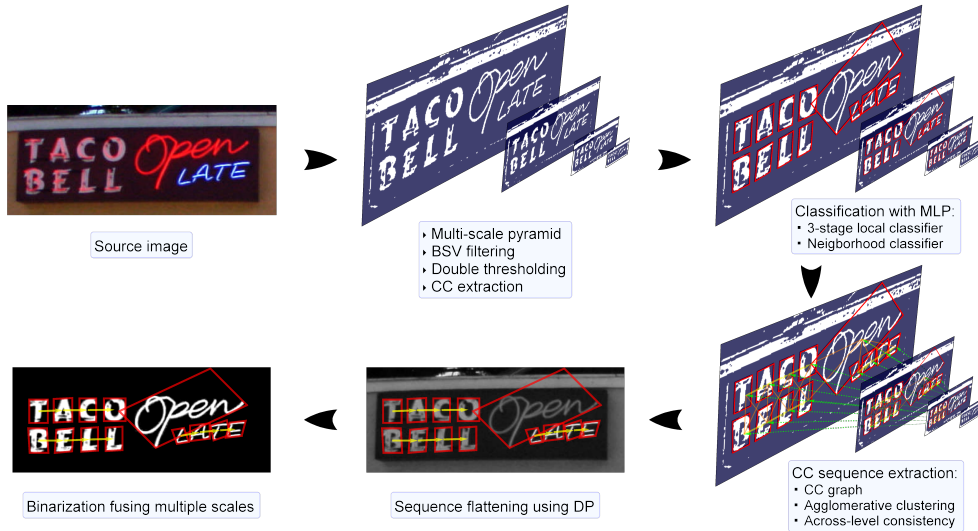
---

* PhD advisor: Carme Julià

Fig. 1. Overview of the proposed text detection and localization system.

methods are accuracy and robustness to noise, but this approach has a high computational cost and most systems can only handle horizontal text.

Methods based on connected components, on the other hand, extract individual connected components (CC) based on local properties such as intensity, color or stroke width. Then, a classifier or some heuristic is used to keep the CC that correspond to text fragments, followed by clustering to text lines. The main advantage of CC based methods is speed: contrary to texture based methods, text at multiple scales and orientations can be detected in few steps, and the number of candidates to be tested can be orders of magnitude smaller than with the sliding window approach. The focus of the majority of the works presented in the literature is classification performance, being the computational cost of the algorithms of secondary importance. However, as many modern portable devices are capable of capturing video at high resolutions (HDV, 4K), there is clearly an opportunity for methods oriented to time efficiency. The method proposed in this paper shares this objective: we present a system capable of real-time localization and segmentation of text in high definition video on a standard laptop, all while keeping a competitive classification performance.

Our method follows the CC-based approach and performs image binarization as a first step of the pipeline (see Fig.1). Our first contribution is a new local binarization method based on a linear spatial filter specially tailored to the "strokeness" property of textual components. Text candidates are extracted by convolving the input image with the filter followed by a thresholding operation. Unlike other local binarization approaches our filter does not require tunable parameters and can detect text comprising a wide range of

4

stroke widths. In addition, to tackle possible variations in the scale of the text beyond the range of the filter we apply it to each level of an image pyramid.

In the second step of the pipeline, a local classifier consisting of a cascade of 3 multilayer perceptron (MLP) is used to filter out non-textual components. The classifier is very fast, taking less than 3 µs on average to check each component. A key contribution in this stage is that we efficiently estimate the stroke width of the components based on the size of their maximal inscribed squares computed using the algorithm described in [5]. As this feature can be obtained inexpensively it can be readily used in the first level of the cascade thus improving its discriminative capacity and the overall processing time.

Running the local classifier is not enough to ascertain certain characters such as 'l' or 'I', which can be easily mistaken for elements in the background. For this reason, an additional classifier that takes into account more global information is run on each candidate CC and its vicinity. A feature vector consisting of individual features of the target CC and its neighbors arranged according to their relative positions and distances is constructed and supplied to a MLP classifier. Our contribution here is the use of the candidate CC returned by the local classifier as seed elements that are later checked with the neighborhood classifier, rather than the more CPU intensive approach of constructing and performing inference on a MRF or CRF used in other methods.

The rest of the paper is organized as follows. Section 2 describes the proposed filter. The experimental evaluation is covered in Section 3. Finally, the article is concluded in Section 4.

2 Proposed Linear Spatial Filter

We propose a new isotropic filter which is suitable for direct CC extraction. We derive our filter from the realization that scene text usually features strong edges at its boundaries while the text interior and its background tend to be more uniform. For a pixel $\mathbf{p}'$ in the text boundary the vector orthogonal to the image gradient at $\mathbf{p}'$ $(\nabla I(\mathbf{p}')^{\perp})$ is approximately tangent to the text outline (see Fig. 2a). If we define $\mathbf{l}$ as the line that passes through $\mathbf{p}'$ with direction $\nabla I(\mathbf{p}')^{\perp}$, pixels in a sufficiently close neighborhood of $\mathbf{p}'$ can be classified as being either interior or exterior depending on which side of $\mathbf{l}$ they lie on. Furthermore, for an arbitrary pixel $\mathbf{p}$ in the image we can accumulate the votes cast by each pixel $\mathbf{p}'$ in the text boundary, weighted by the intensity of $\nabla I(\mathbf{p}')$ and the distance from $\mathbf{p}$ to $\mathbf{p}'$ (see Fig. 2b). The sign of each vote can be determined by examining which side of $\mathbf{l}$ $\mathbf{p}$ lies on. If the accumulated result is greater (or less, depending on the polarity of the text) than a threshold then $\mathbf{p}$ is interior to the text outline.

We can make an analogy with physics and regard the perimeter of the text as a closed electrical circuit. In this case the set of vectors $\nabla I(\mathbf{p}')^{\perp}$
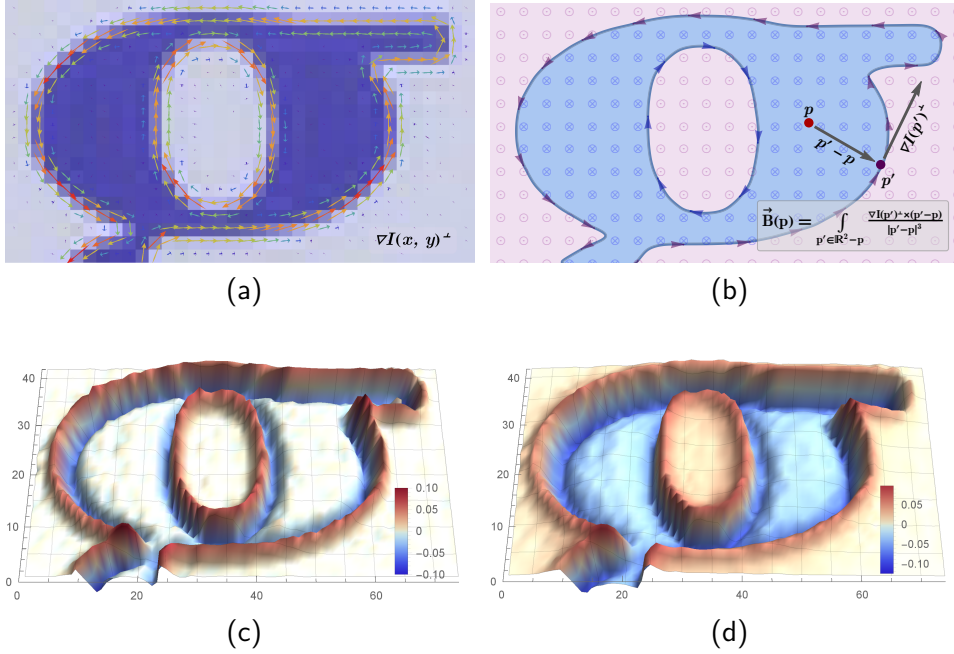
Fig. 2. (a) The vector orthogonal to the image gradient $\nabla I(x,y)^{\perp}$ is tangent to the text outline. (b) Each pixel accumulates votes from its neighbors according to (1). (c) Result of convolving the image in (a) with a L$_1$-normalized LoG with $\sigma=0.91$. (d) Result of convolving the image in (a) with the proposed filter kernel (10).

are akin to differential elements of electric current circulating around the text boundary and the induced magnetic field $\vec{B}$ will follow the Biot-Savart Law[2,3] (see Fig. 2b).

$$\vec{B}(\mathbf{p}) = \int\limits_{\mathbf{p}'\in\mathbb{R}^2-\mathbf{p}} \frac{\nabla I(\mathbf{p}')^{\perp}\times(\mathbf{p}'-\mathbf{p})}{|\mathbf{p}'-\mathbf{p}|^3}d\mathbf{p}' \tag{1}$$

Since the circuit is planar, $\vec{B}$ has zero x and y components in the plane of the image while the value of the z component is

$$\mathrm{B}(x,y) = \int\limits_{\substack{u,v\in\mathbb{R}\\u,v\neq x,y}} \frac{\nabla I_x(u,v)(x-u)+\nabla I_y(u,v)(y-v)}{\left((x-u)^2+(y-v)^2\right)^{3/2}}du\,dv \tag{2}$$

It can be shown that (2) can be expressed as the sum of the convolutions of the x and y components of the image gradient with two linear translation-invariant filters, respectively.

---

[2] Physics constants have been removed.

[3] 2D vectors in the cross product extended to 3D by setting z to 0.

$$B(x, y) = [\nabla I_x * h_x + \nabla I_y * h_y](x, y) \tag{3}$$

The filter kernels are

$$h_x(x, y) = \begin{cases} 0 & x = 0 \wedge y = 0 \\ \frac{-x}{(x^2+y^2)^{3/2}} & \text{otherwise} \end{cases}$$
$$h_y(x, y) = \begin{cases} 0 & x = 0 \wedge y = 0 \\ \frac{-y}{(x^2+y^2)^{3/2}} & \text{otherwise} \end{cases} \tag{4}$$

The image gradients, in turn, can be estimated by convolving the image with the Gaussian derivatives

$$\nabla I_x = I * \text{DoG}_x \quad \nabla I_y = I * \text{DoG}_y \tag{5}$$

$$\text{DoG}_x(x, y; \sigma) = -\frac{x}{2\pi\sigma^4} e^{-\frac{x^2+y^2}{2\sigma^2}}$$
$$\text{DoG}_y(x, y; \sigma) = -\frac{y}{2\pi\sigma^4} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{6}$$

Grouping all terms and applying the associativity and common factor properties of the convolution, (3) can be reduced to a convolution of the image with a single filter kernel

$$B(x, y) = [I * \underbrace{(\text{DoG}_x * h_x + \text{DoG}_y * h_y)}_{\text{bsv}(x,y)}](x, y) \tag{7}$$

In the discrete domain, the limit of the normalized DoG kernels when $\sigma \to 0$ are the centered difference kernels, which when substituted into (7) result in

$$\text{bsv}(x, y) = [1\ 0\ {-1}] * h_x + [1\ 0\ {-1}]^{\text{T}} * h_y \tag{8}$$

For practical uses, the discrete filter kernel in (8) must be truncated to a finite window size W. In addition, we normalize the truncated kernel to make it zero mean and its L_1 entrywise norm equal to 1.

$$\overset{\wedge}{\text{BSV}} = [\text{bsv}(j - \tfrac{W-1}{2}, i - \tfrac{W-1}{2})]_{W \times W} \tag{9}$$

$$\text{BSV} = \frac{\overset{\wedge}{\text{BSV}} - \overline{\text{BSV}}}{\left\| \overset{\wedge}{\text{BSV}} - \overline{\text{BSV}} \right\|_1} \tag{10}$$

The filter in (10) is similar to a LoG albeit with a slower decay rate stemming from the inverse-square attenuation term in the Biot-Savart Law (1).

Fig. 3. Comparison of local binarization methods. (a) Source image, (b) Niblack [11], (c) Bernsen [1], (d) Sauvola and Pietikinen [13], (e) Wolf *et al.* [15], (f) Bradley and Roth [2], (g) NICK [7], (h) Proposed method.

3 Experimental results

We begin by comparing the proposed component segmentation scheme using the filter described in Section 2 with 6 well known local binarization methods for which $\mathcal{O}(n)$ time complexity implementations exist. In order to make a fair comparison all the evaluated methods have been applied in a multi-scale fashion: First, the source image is converted to grayscale and a dyadic pyramid is generated. Then, all levels of the pyramid are binarized using the methods being compared in both dark on light and light on dark polarities, and CC are extracted; in all cases, the threshold and smoothing parameters of the algorithms are set to the values suggested by the authors, and the window size is empirically set to 15. Next, discrete optimization is employed to pick the set of disjoint CC from all pyramid levels and polarities that best matches the pixel-level ground truth (GT). Finally, the pixel-wise precision, recall, and F-score measures are computed from the set of selected CC and the GT segmentation.

For this comparison, two datasets that provide pixel-level GT are considered: the ICDAR 2013 Robust Reading Competition Challenge 2 dataset [6], and the KAIST dataset [8]. The ICDAR 2013 dataset consists of 229 training images and 233 images for the test set, which are all employed since there are no classifiers involved in this stage. As regards the KAIST dataset, the entire set of 2483 images are used.

The quantitative results of the evaluation are presented in Table 1, while some qualitative results are shown in Fig. 3. The proposed method attains the best results in terms of recall and average F-score on both datasets. In particular, the recall rate on the ICDAR 2013 dataset is 0.89, which is a 3%

Table 1. Comparison of multi-scale implementations of binarization methods on the ICDAR 2013 [6] and KAIST [8] datasets.

Method	ICDAR 2013 dataset			KAIST dataset		
	p	r	f	p	r	f
Niblack [11]	0.93	0.86	0.90	0.83	0.97	0.89
Bernsen [1]	0.91	0.86	0.89	0.80	0.97	0.87
Sauvola and Pietikinen [13]	0.96	0.75	0.82	0.87	0.91	0.87
Wolf *et al.* [15]	0.97	0.70	0.81	0.88	0.87	0.85
Bradley and Roth [2]	0.95	0.82	0.87	0.85	0.96	0.90
NICK [7]	**0.97**	0.72	0.82	**0.90**	0.91	0.89
Proposed	0.93	**0.89**	**0.91**	0.84	**0.98**	**0.90**

The best results in each category are highlighted in **bold**.

The f columns show averaged F-score values.

higher than the ones achieved by Niblack [11] and Bernsen [1]. In the case of the KAIST dataset the results are much closer: a recall of 0.98 for the proposed method compared to the 0.97 recall of the Niblack and Bernsen algorithms. On the other hand, the precision obtained with all methods is low relative to the recall rates, which could be explained by the fact that the GT in this dataset is noticeably thinner than the corresponding text in the source images.

Next, we evaluate the proposed text localization pipeline on the 233 test images of the ICDAR 2013 dataset using the metric described in Wolf et al. [16] and the DetEval tool. The results of the evaluation are summarized in Table 2.

The proposed method achieves a F-score of 0.790, which is competitive with the state-of-the-art for non deep learning based methods. It also attains the highest recall rate (0.794) partly due to its capability to detect non-horizonal and curved text lines (see Fig. 4 for some examples), and also because all non-singleton sequences are accepted for evaluation. The downside is an increased number of false positives, which results in the second lowest precision (0.787) of all the compared methods. In the category of processing speed, our method performs more than 2 times faster than the second best method, and more than 5 times faster if we employ a parallel implementation. All experiments have been carried out on an Intel® Core™ i7-3840QM 2.80 GHz laptop.

4 Conclusion

A text localization method in natural scene images based on the CC approach is presented in this paper. The method introduces a number of innovations. First, a specifically designed linear spatial filter is used for image binarization.

Table 2. Comparison of text localization results on the ICDAR 2013 dataset [6].

Method	p	r	f	t (ms)	CPU (GHz)
TexStar (ICDAR'13 winner) [6]	0.885	0.664	0.759		
Zamberletti *et al.* [17]	0.86	0.70	0.77	750[1]	
FASText [3]	0.840	0.693	0.768	150	2.40
Neumann and Matas [10]	0.818	0.724	0.771	400	2.70
Lu *et al.* [9]	**0.892**	0.686	0.782	10000[2]	
Zhang *et al.* [18]	0.88	0.74	0.80	30000[2,3]	2.00
Text Flow [14]	0.851	0.759	0.802	1500[3]	2.00
Text Catcher [4]	0.755	0.756	0.755	6000	3.10
Zhu *et al.* [19]	0.85	0.75	0.79		
Qin and Manduchi (fast) [12]	0.846	0.752	0.796	380[1]	3.40
Qin and Manduchi [12]	0.888	0.787	**0.834**	1200[1]	3.40
Proposed	0.787	**0.794**	0.790	**68**[4]	2.80

The best results in each category are highlighted in **bold**.

Processing times correspond to serial implementations of the methods and have been measured on the ICDAR'13 dataset, unless specified otherwise.

[1] Parallel implementation. Processing times measured on 640x480 images.
[2] Implemented in MATLAB.
[3] Processing times measured on the ICDAR'11 dataset.
[4] The average processing time for the parallel implementation is 29 ms.

Fig. 4. Examples of text localization using the proposed method.

The filter has similar properties to the Laplacian of Gaussian albeit with a slower attenuation, which makes it suitable for direct CC extraction. Second, the stroke width of a CC is approximated as the average of the sides of the maximal inscribed squares obtained with the medial axis transform (MAT), which is computed in linear time employing the algorithm described in [5]. Last, a MLP is employed to classify CC considering their local neighborhood, which is faster than building a dense CRF and performing inference on it.

The proposed method has been evaluated on the ICDAR 2013 Robust Reading Competition Challenge 2 dataset [6] using the metric in [16], achieving F-score of 0.790, which is competitive with the state-of-the-art. In terms of processing speed, our method attained the highest performance, being more than 2 times faster than the second best method.

The main contribution of the proposed method is its execution speed, being capable of processing 1080p HD video at nearly 30 frames per second on a standard laptop.

References

[1] John Bernsen. Dynamic Thresholding of Grey-Level Images. In *ICPR*, volume 2, pages 1251–1255, 1986.

[2] Derek Bradley and Gerhard Roth. G.: Adaptive Thresholding Using the Integral Image. *Journal of Graphics Tools*, pages 13–21, 2007.

[3] M. Buta, L. Neumann, and J. Matas. FASText: Efficient Unconstrained Scene Text Detector. In *ICCV*, pages 1206–1214, December 2015.

[4] Jonathan Fabrizio, Myriam Robert-Seidowsky, Séverine Dubuisson, Stefania Calarasanu, and Raphaël Boissel. TextCatcher: A method to detect curved and challenging text in natural scenes. *IJDAR*, 19(2):99–117, March 2016.

[5] J.F. Jenq and S. Sahni. Serial and Parallel Algorithms for the Medial Axis Transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(12):1218 –1224, December 1992.

[6] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazàn, and L. P. de las Heras. ICDAR 2013 Robust Reading Competition. In *ICDAR*, pages 1484–1493, August 2013.

[7] Khurram Khurshid, Imran Siddiqi, Claudie Faure, and Nicole Vincent. Comparison of Niblack inspired binarization methods for ancient documents. In *SPIE*, volume 7247, pages 72470U–72470U–9, 2009.

[8] SeongHun Lee, Min Su Cho, Kyomin Jung, and Jin Hyung Kim. Scene Text Extraction with Edge Constraint and Text Collinearity. In *ICPR*, pages 3983–3986, 2010.

[9] Shijian Lu, Tao Chen, Shangxuan Tian, Joo-Hwee Lim, and Chew-Lim Tan. Scene Text Extraction Based on Edges and Support Vector Regression. *IJDAR*, 18(2):125–135, February 2015.

[10] L. Neumann and J. Matas. Efficient Scene Text Localization and Recognition with Local Character Refinement. In *ICDAR*, pages 746–750, August 2015.

[11] Wayne Niblack. *An Introduction to Digital Image Processing*. Prentice Hall, First English Edition edition, July 1986.

[12] S. Qin and R. Manduchi. A Fast and Robust Text Spotter. In *WACV*, pages 1–8, March 2016.

[13] J. Sauvola and M. Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, February 2000.

[14] S. Tian, Y. Pan, C. Huang, S. Lu, K. Yu, and C. L. Tan. Text Flow: A Unified Text Detection System in Natural Scene Images. In *ICCV*, pages 4651–4659, December 2015.

[15] C. Wolf, J. M. Jolion, and F. Chassaing. Text Localization, Enhancement and Binarization in Multimedia Documents. In *ICPR*, volume 2, pages 1037–1040 vol.2, 2002.

[16] Christian Wolf and Jean Michel Jolion. Object Count/Area Graphs for the Evaluation of Object Detection and Segmentation Algorithms. *IJDAR*, 8(4):280–296, April 2006.

[17] Alessandro Zamberletti, Lucia Noce, and Ignazio Gallo. Text Localization Based on Fast Feature Pyramids and Multi-Resolution Maximally Stable Extremal Regions. In *ACCV Workshops*, number 9009, pages 91–105. November 2014.

[18] Zheng Zhang, Wei Shen, Cong Yao, and Xiang Bai. Symmetry-Based Text Line Detection in Natural Scenes. In *CVPR*, pages 2558–2567, June 2015.

[19] R. Zhu, X. J. Mao, Q. H. Zhu, N. Li, and Y. B. Yang. Text Detection Based on Convolutional Neural Networks with Spatial Pyramid Pooling. In *ICIP*, pages 1032–1036, September 2016.

On-line method to deduce the costs of graph edit distance for handwritten character recognition

Elena Rica [*]

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
`mariaelena.rica@urv.cat`

This paper is twofold, on the one hand, we list several methods that have been presented to deduce the insertion, deletion and substitution costs of the graph edit distance and, on the other hand, we present an on-line learning method to automatically deduce them. It is inspired in a previously published off-line learning method based on embedding the ground-truth node-to-node mappings into a Euclidean space and learning the edit costs through the hyper-plane that splits the nodes into the mapped and the non-mapped ones. The new method has the advantage of learning the edit costs and computing the graph edit distance can be done simultaneously. Experimental validation shows that the matching accuracy is competitive with the off-line method but without the need of the whole learning set.

Keywords: Graph-matching algorithm, Graph edit distance, Learning edit costs, On-line learning.

1 Introduction

Attributed relational graphs are commonly used as abstract representations for common structures such as documents, images or chemical compounds, among others [23]. Nodes of graphs represent local parts of the object and edges represent the relations between these local parts. Error-tolerant graph matching algorithms [4,31] are applied to deduce the distance between prototypes represented by attributed graphs. Error-tolerant graph matching algorithms are based on finding a mapping between nodes so that both graphs look similar when their nodes are mapped according to this node-to-node mapping. One of the most used frameworks to define the error-tolerant graph matching is through the graph edit distance [30,11,28]. The main idea is to define the difference between graphs as the amount of distortion required to transform one graph into another through substituting, deleting or inserting nodes and edges. To do so, some penalty costs are defined for these edit operations.

In this paper, we list several methods that have been presented to deduce these costs and we present an on-line method to learn them. The aim is to recompute these

---

[*] PhD advisors: Susana Álvarez, Francesc Serratosa

values automatically when new data is available. Thus, in a recognition process, the graph edit distance can be computed having these costs as input parameters, which have been found trough our optimisation process. In this way, the recognition process can be carried out at the same time than the learning process. The method is inspired by an off-line learning algorithm recently published [1]. Nevertheless, our algorithm is not a simple iterating process on the off-line algorithm, but some processes have been incorporated to keep the algorithm learning with the minimum data.

Note that, in some object retrieval applications, in which elements are represented by graphs, the aim is to deduce which are the most similar graphs, without the graphs being previously classified. In these cases, it is crucial to learn the edit costs such that the best node-to-node mapping between pairs of graphs is computed instead of maximising the classification ratio. This is the reason why the whole process we present is dependent on a ground-truth node-to-node matching. Recently, a graph database generator that returns pairs of graphs with their ground-truth correspondence has been presented [27].

Table 1 shows several off-line published methods. An important feature of them is the type of costs the learning algorithm obtains: a self-organising map [19], a probability density function [20] or linear functions [1,29,2,7,15,6]. These off-line methods learn with the whole data at once, however, the on-line methods receive subsets of data and make successive learning processes with them. To our knowledge, any on-line method has been published yet to learn the graph edit distance parameters and we present one that learns the insertion and deletions costs (similarly to [6,1]) and the weights on the substitutions costs (similarly to [2,7,15,1]).

Table 1: Published methods related on learning the edit costs.

Ref.	Authors	Objective Function	Learning method
2005 [19]	Neuhaus Bunke	Average of 8 indices: DaviesBouldin [8], Dunn [9], C [13], GoodmanKrusk [12], CalinskiHaraba [3], Rand [21], Jaccard [14], FowlkesMallo [10],	The method learns the weights of a Self Organized Map (SOM) to define the substitution, deletion and insertion costs on nodes and edges. These costs become the output of the SOM when the input is the attribute of the node or the edge. **Learns:** F_S^n, F_D^n, F_I^n, F_S^e, F_D^e, F_I^e.
2007 [20]	Neuhaus Bunke	Dunn Index [22]	The method learns the parameters of a Probability Density Function (PDF) to define the substitution, deletion and insertion costs on nodes and edges. These costs become the inverse of the probability set by the PDF given the attributes of the node or the edge. **Learns:** F_S^n, F_D^n, F_I^n, F_S^e, F_D^e, F_I^e.

Continued on next page

14

Table 1 – *Continued from previous page*

Ref.	Authors	Objective Function	Learning method
2009 [**2**]	Caetano McAuley Cheng Le Smola	Correspondence accuracy	The method learns the weights of the weighted Euclidean distance to define the substitution cost on nodes and edges. The substitution cost becomes the weighted Euclidean distance for nodes and edges. Insertion and deletion of nodes and edges are not learned and assumed to be constant. **Learns:** w_S^n, w_S^e.
2012 [**15**]	Leordeanu Sukthankar Hebert	Recognition ratio	The same than [2].
2015 [**6**]	Cortés Serratosa	Correspondence accuracy	The method learns the deletion and insertion costs on nodes and edges as constants (Real numbers). The substitution cost is assumed the Euclidean distance between the attributes on nodes or on edges. **Learns:** K_D^n, K_I^n, K_D^e, K_I^e.
2016 [**7**]	Cortés Serratosa	Correspondence accuracy	The same than [2].
2017 [**22**]	Raveaux Martineau Conte Venturini	Recognition ratio	The same than [2].
2018 [**5**]	Cortés Conte Cardot Serratosa	Correspondence accuracy	The method learns the substitution functions on nodes and edges through a Neural Network (NN). The substitution cost is defined as the output of the NN when the input is the attribute on the nodes and edges. Insertion and deletion of nodes and edges are not learned and assumed to be constants. **Learns:** F_S^n, F_S^e,
2018 [**24**]	Santacruz Serratosa	Correspondence accuracy	Similar to [5] but the insertion and deletion costs on nodes and edges are also learned. There is also a NN for insertion and another one for deletion the nodes and edges. **Learns:** F_S^n, F_D^n, F_I^n, F_S^e, F_D^e, F_I^e.
2018 [**1**]	Algabli Serratosa	Correspondence accuracy	The method learns the weights of the weighted Euclidean distance to define the substitution cost and also the deletion and insertion costs as constants on nodes and edges. The substitution cost is computed as a weighted Euclidean distance in which the weights have been learned. The insertion and deletion costs become the learned constant (Real number). **Learns:** w_S^n, K_D^n, K_I^n, w_S^e, K_D^e, K_I^e.

Table 1 – *Continued from previous page*

Ref.	Authors	Objective Function	Learning method
2018 [**17**]	Martineau Raveaux Conte Venturini	Recognition ratio	The method learns the weights on each node or edge. These weights depend on how important are the nodes and edges to describe the class. **Learns:** weights on nodes and edges.

The outline of the paper is as follows. In the next section, we define attributed graphs and graph edit distance. In section 3, we present our learning strategy. In section 4, we show the experimental validation and finally, in section 5, we conclude the article.

2 Graph Edit Distance

The graph edit distance (GED) between two attributed graphs is defined as the transformation from one graph into another, through the edit operations, that obtains the minimum cost. These edit operations are: substitution, deletion and insertion of nodes and also edges. Every edit operation has a cost depending on the attributes on the involved nodes or edges. This graph transformation can be defined through a node-to-node mapping f between nodes of both graphs.

Having a pair of graphs, G and G', a correspondence f between these graphs is a bijective function that assigns one node of G to only one node of G'. We suppose both graphs have the same number of nodes since they have been expanded with new nodes that have a concrete attribute. We call these new nodes as $Null$. Note that the mapping between edges is imposed by the mapping of the nodes whose edges are connected.

We define G_i as the ith node in G, G'_a as the ath node in G', $G_{i,j}$ as the edge between the ith node and the jth node in G, $G'_{a,b}$ as the edge between the ath node and the bth node in G'. Nodes and edges have N and M attributes, which are real numbers, respectively. Moreover, γ_i^t is the tth attribute of node G_i and $\beta_{i,j}^t$ is the tth attribute of edge $G_{i,j}$. We also define the mapping $f(i) = a$ from G_i to G'_a, we say that it represents a node substitution if both nodes are not $Null$. Contrarily, if node G'_a is a $Null$ and G_i is not, we say that it represents a deletion. Finally, if node G_i is a $Null$ and G'_a is not, we say that it represents an insertion. Similarly happens with the edges. The case that both nodes or both edges are null is not considered since it is defined as the cost is always zero.

We define the GED as follows:

$$GED(G, G') = \min_{f:G \to G'} \left\{ \sum_{\forall G_i} C^n(i, f(i)) + \sum_{\forall G_{i,j}} C^e(i, j, f(i), f(j)) \right\} \qquad (1)$$

Where, functions $C^n(i, f(i))$ and $C^e(i, j, f(i), f(j))$ represent the cost of mapping a pair of nodes (G_i and $G'_{f(i)}$) and a pair of edges ($G_{i,j}$ and $G'_{f(i),f(j)}$), respectively, and they are defined through the cost functions in Eq. 2 and Eq. 3.

16

$$C^n(i,a) = \begin{cases} C^n_S(i,a) & \text{if } G_i \neq Null \wedge G'_a \neq Null \\ C^n_D(i) & \text{if } G_i \neq Null \wedge G'_a = Null \\ C^n_I(a) & \text{if } G_i = Null \wedge G'_a \neq Null \end{cases} \qquad (2)$$

$$C^e(i,j,a,b) = \begin{cases} C^e_S(i,j,a,b) & \text{if } G_{i,j} \neq Null \wedge G'_{a,b} \neq Null \\ C^e_D(i,j) & \text{if } G_{i,j} \neq Null \wedge G'_{a,b} = Null \\ C^e_I(a,b) & \text{if } G_{i,j} = Null \wedge G'_{a,b} \neq Null \end{cases} \qquad (3)$$

Where $C^n_S(i,a)$ is the cost of substituting node G_i by node G'_a, $C^n_D(i)$ is the cost of deleting node G_i and $C^n_I(a)$ is the cost of inserting node G'_a. Similarly, $C^e_S(i,j,a,b)$ is the cost of substituting edge $G_{i,j}$ by edge $G'_{a,b}$, $C^e_D(i,j)$ is the cost of deleting edge $G_{i,j}$ and $C^e_I(a,b)$ is the cost of inserting edge $G'_{a,b}$.

In this paper, we impose the restrictions $C^n_I(a) = C^n_D(i) = K^n$ and $C^e_I(i,j) = C^e_D(i,j) = K^e$, where K^n and K^e are real numbers. Moreover,

$C^n_S(i,a) = \sum_{t=1}^{N} w^n_t \left\| \gamma^t_i - \gamma'^t_a \right\|$ and $C^e_S(i,j,a,b) = \sum_{t=1}^{M} w^e_t \left\| \beta^t_{i,j} - \beta'^t_{a,b} \right\|$, where $w^n = (w^n_1, ..., w^n_N)$ is the vector of nodes attributes' weights and $w^e = (w^e_1, ..., w^e_M)$ is the vector of edges attributes' weights. Furthermore,

$$1 = \sum_{t=1}^{N} w^n_t \qquad 1 = \sum_{t=1}^{M} w^e_t \qquad (4)$$

3 Learning the Graph edit costs

In this section, we first plainly summarise the off-line method presented in [1] and then we move on to explain our on-line proposal. It is crucial to explain the off-line method since our method is inspired in it.

3.1 Off-line learning the graph edit costs

The basic scheme of the off-line method is summarised in Figure 1. The system receives a set of triplets composed of two graphs and a ground-truth correspondence between them, $\{(G, G', f)_1, (G, G', f)_2, ...\}$, and outputs the substitution weights on nodes and edges and also the deletion and insertion costs on nodes and edges. Figure 1 only shows one triplet composed of two graphs that have five and four nodes, respectively. The ground-truth correspondence is represented through the dashed arrows. Four nodes are substituted and one node is deleted.

The algorithm is composed of three main steps:

In the first step (Embedding), the ground-truth node-to-node mappings are embedded into a Euclidean space S, being $S = (S^n_1, ..., S^n_N, S^e_1, ..., S^e_M, S_{K^e})$ of dimension $N + M + 1$. Each node substitution is transformed into a point in this space and it is assigned to the "$+1$" class. Moreover, each node deletion is transformed into $\tilde{N}$ points, which are assigned to the "-1" class. $\tilde{N}$ is the number of substituted nodes in the ground-truth correspondence. The ground-truth correspondence in Figure 1 makes the embedding step to generate four points that represent the four node substitution operations (one point per substitution) and four points that represent the only one node deletion (the number of points that generate each deletion is the number of substituted nodes).
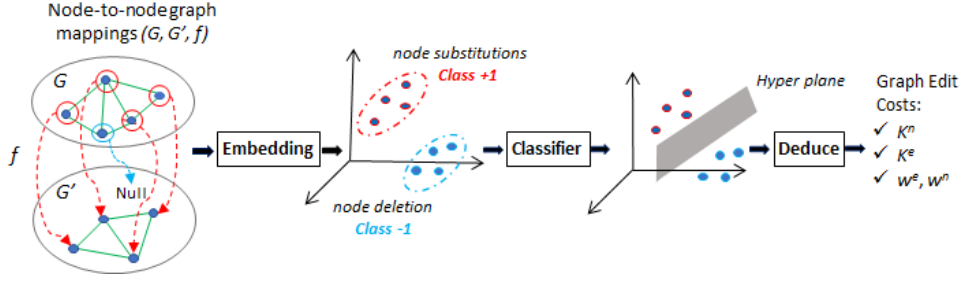
Fig. 1. Basic scheme of the off-line learning method.

In the second step (Classifier), a linear hyper-plane is computed that has to be the best linear border between both classes. Authors in [1] describe that any known linear classifier that return the hyper-plane can be used. Equation 5 defines this border, as described in [1]. Note the constants in this hyper-plane are the substitution weights $w_2^n, ..., w_N^n$ and $w_2^e, ..., w_M^e$ and also the insertion and deletion costs on nodes and edges K^n and K^e, respectively. Finally, note that w_1^n and w_1^e do not appear in Equation 5.

$$S_1^n + w_2^n \cdot S_2^n + ... + w_N^n \cdot S_N^n + S_1^e + w_2^e \cdot S_2^e + ... + w_M^e \cdot S_M^e + \\ K^e \cdot S_{K^e} + K^n = 0 \tag{5}$$

For explanatory reasons, Figure 2 shows the specific case of $N = M = 1$, where S is a 3D dimensional space. In this example, graphs have three and two nodes (not shown in the figure). The ground-truth correspondence imposes two nodes to be substituted (they generate two points) and one node to be deleted (that also generates two points).
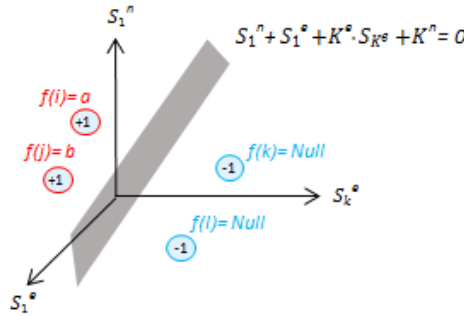

Fig. 2. Hyper-plane obtained in the learning process when M=N=1.

Finally, in the last step (Deduce), weights $w_2^n, ..., w_N^n$ and $w_2^e, ..., w_M^e$ and also constants K^n, and K^e are extracted from the hyper-plane constants. Moreover, w_1^n and w_1^e are obtained through Equation 4.

3.2 On-line learning the graph edit costs

While the off-line method embeds all the set of triplets at once, the on-line method receives a triplet (G, G', f) at a time and computes K^n, K^e, w^n and w^e each time it is executed. To do so, it is needed the algorithm to input not only (G, G', f) but also two sets of embedded points D_{-1} and D_1 that have been computed in the previous execution (they are empty sets in the first iteration of the algorithm). The algorithm is composed of four main steps (Figure 3):
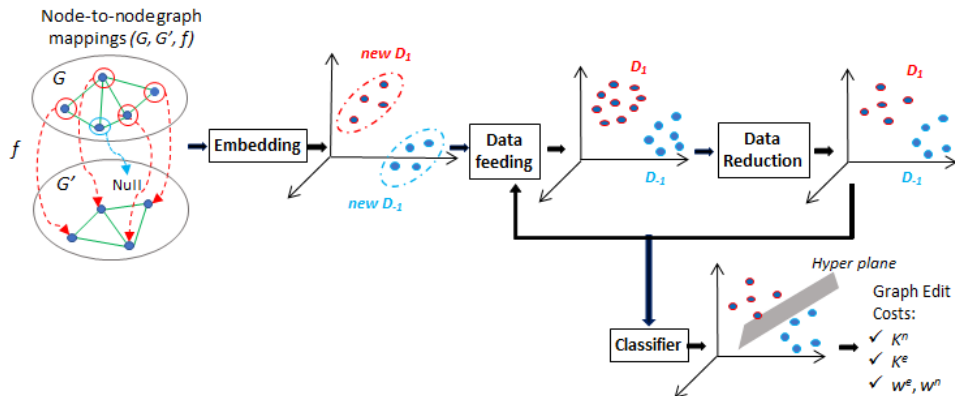

Fig. 3. Basic scheme of the on-line learning method.

Embedding (G, G', f) into the Euclidean space S is the first step of the algorithm (Line 1) and it is done in the same way that the off-line method [1] does. More precisely, it generates two sets, *new* D_1 and *new* D_{-1}, composed of points in the space S that represent the node substitutions and node deletions in (G, G', f), respectively.

Feeding (Lines 2 and 3) is a simple process in which the new sets *new* D_1 and *new* D_{-1} and the previous ones D_1 and D_{-1} (which are input parameters of the algorithm) are put together.

Then, Data reduction (Lines 4 to 14) updates sets D_1 and D_{-1} with the aim of reducing the amount of points but holding two main properties of these sets. The first one is keeping the general distance between points as well as their positions. This means that we want to have less points but maintain the same information of the sets as much as possible. The second property is keeping the same relation of the number of points of both sets. This is because, all the classifiers are biased by the order of the sets. In this case, we want to keep the number of points proportion to be as much reliable as possible to the input data. The input parameter k^{means} is the maximum number of points that will have the sets when each iterations finishes and the algorithm returns the graph edit distance parameters. From Line 4 to Line 11, the algorithm decides the number of elements that the updated sets D_1 and D_{-1} will have. Finally, in Lines 12 and 13, the reduction is done in each set. Note that we have selected the K-means clustering method [16] to perform this reduction although other reduction algorithms could be explored. Note the generated sets D_1 and D_{-1} are returned by the algorithm and feed the next iteration.

Finally, in the Classifier step (Lines 15 and 16), $w_2^n, ..., w_N^n, w_2^e, ..., w_M^e, K^n$ and K^e are extracted from the hyper-plane constants that a linear classifier returns as the border of sets D_1 and D_{-1}. Moreover, w_1^n and w_1^e are obtained through Equation 4.

On-line Algorithm

Input(K^{means}, D_1, D_{-1}, (G, G', f)) $\rightarrow$ Output(K^n, K^e, w^n, w^e, D_1, D_{-1})

1. (*new D_1*, *new D_{-1}*) = Embedding(G, G', f)

2. $D_1 = D_1 \cup new \ D_1$

3. $D_{-1} = D_{-1} \cup new \ D_{-1}$

4. **If** $|D_1| > K^{means} \vee |D_{-1}| > K^{means}$

5.   **If** $|D_1| \geq |D_{-1}|$

6. $K_1^{means} = K^{means}$

7. $K_{-1}^{means} = K^{means} * (|D_{-1}| / |D_1|)$

8.   **Else**

9. $K_1^{means} = K^{means} * (|D_1| / |D_{-1}|)$

10. $K_{-1}^{means} = K^{means}$

11.   **End if**

12. $D_1 = k\text{-}means(D_1, K_1^{means})$

13. $D_{-1} = k\text{-}means(D_{-1}, K_{-1}^{means})$

14. **End if**

15. $[K^n, K^e, w_2^n, ..., w_N^n, w_2^e, ..., w_M^e]$ = Classifier (D_1, D_{-1})

16. $w_1^n = 1 - \sum_{t=2}^{N} w_t^n \quad w_1^e = 1 - \sum_{t=2}^{M} w_t^e$

End Algorithm

4 Experimental Validation

The method we present has been tested using the Tarragona-Graph repository detailed in [18]. This repository has the main characteristic that each register is composed of a pair of graphs, a ground-truth correspondence between them (mapping between their nodes), and their class. It contains three graph databases: Letter_Low, Letter_Med and Letter_High that represent artificially distorted letters of the Latin alphabet with an increasing level of distortion. In each data base we have used a set of $37,500$ pairs of graphs for learning and a different set of $37,500$ pairs of graphs for testing. Every set generated $150,000$ points in the embedding space. We used the matching algorithm [25,26].

The average matching accuracy obtained in the three data sets is shown in Figure 4, Figure 5 and Figure 6, given a number of introduced triplets (G, G', f) taken from the test set and also different values of K^{means}. The value $K^{means} = Inf$ represents no reduction of the data, which is the same than applying the off-line algorithm [1], given the specific number of introduced triplets.
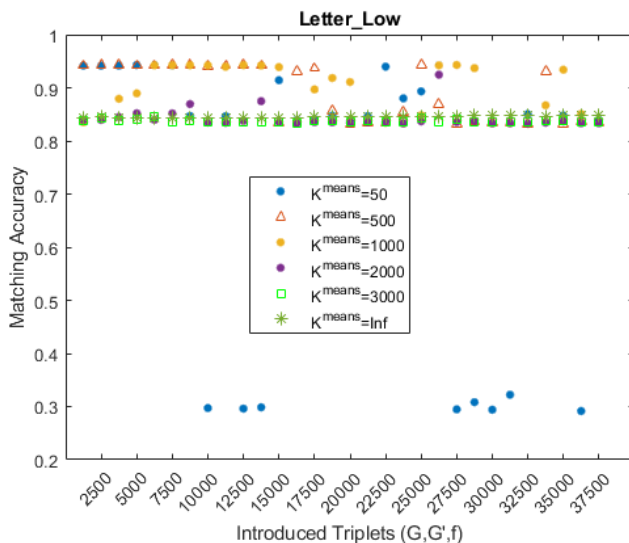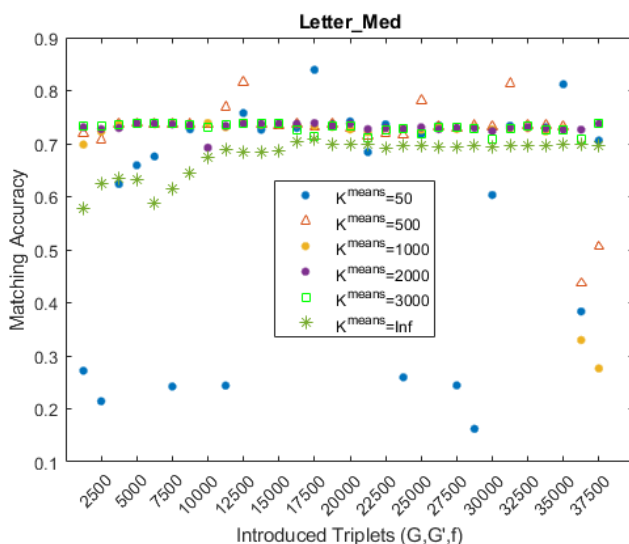
Fig. 4. Letter_Low Accuracy.

Fig. 5. Letter_Med Accuracy.

In Letter_Low (Figure 4), accuracies generated by different values of K^{means} are stable and almost similar except for the ones generated by $K^{means} = 50$. Nevertheless, in Letter_Med (Figure 5) and Letter_High (Figure 6), the stability is achieved at $K^{means} = 2,000$ and $K^{means} = 3,000$, respectively. The off-line method ($K^{means} = Inf$) becomes to be the most stable. Note that, higher is the number of points K^{means} allowed in sets D_1 and D_{-1}, slower is the algorithm (see Table 2). Thus, we wish to keep this value as lower as possible. Nevertheless, we observe that this is a parameter that depends on the level of noise of the databases. Comparing our method to the off-line one [1], we realise that we achieve competitive accuracies, although having

21

a huge data reduction. Note that in these databases, the off-line method generates 150,000 points but the on-line algorithm only needs 3,000 points, in the worst case (Letter_High).

Table 2. Run time in seconds given several $\mathbf{K^{means}}$

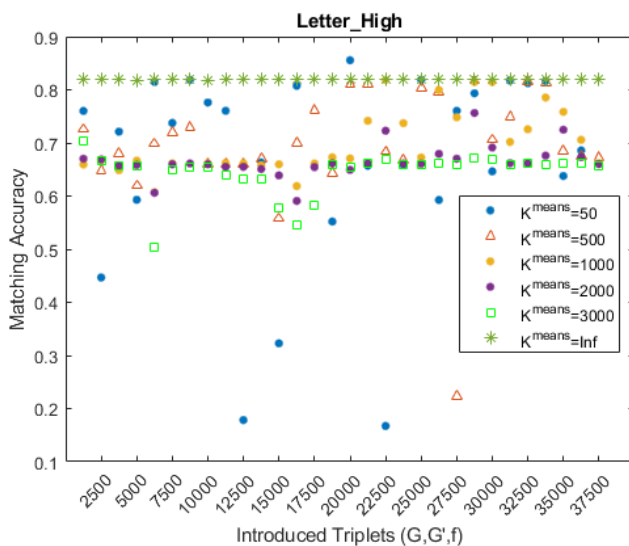| $\mathbf{K^{means}}$	50	500	1000	2000	3000	Off-line
Letter_Low	1.3	6.5	11.7	29.8	52.3	51.7
Letter_Med	1.3	5.5	11.1	25.6	68.4	52.1
Letter_High	1.2	6.6	12.0	25.7	67.1	55.4

Fig. 6. Letter_High Accuracy.

5 Conclusions

We have presented an on-line method to learn the edit costs based on embedding the ground-truth correspondence into a Euclidean space. This space, which was previously defined in an off-line method, has the particularity that the border between node substitutions and deletions is set as a hyper-plane defined by the edit cost parameters. The learning method is limited to the applications that substitution costs are represented as a normalised euclidean distance and insertion and deletion costs are constants. Note that the weights and costs deduced through our algorithm do not guarantee to be the optimal ones in an optimal graph-matching algorithm. Each time our method is executed, the weights and edit costs are returned and also some embedded points. These points and the new node-to-node mappings are the input of

the next algorithm iteration. From a practical point of view, our method has three main advantages. First, the learned weights and costs can be used each time the algorithm is computed. Second, only parameter K^{means} has to be tuned. And third, the graph edit distance does not need to be computed in the learning process, as it is needed in other methods. Moreover, the experimental validation shows the learned parameters obtain an accuracy that is similar to the off-line method in few iterations and having an important data reduction.

References

[1] Algabli, S., Serratosa, F.: Embedding the node-to-node mappings to learn the graph edit distance parameters. Pattern Recognition Letters **112**, 353–360 (2018)

[2] Caetano, T.S., McAuley, J.J., Cheng, L., Le, Q.V., Smola, A.J.: Learning graph matching. IEEE transactions on pattern analysis and machine intelligence **31**(6), 1048–1058 (2009)

[3] Caliński, T., Harabasz, J.: A dendrite method for cluster analysis. Communications in Statistics-theory and Methods **3**(1), 1–27 (1974)

[4] Conte, D., Foggia, P., Sansone, C., Vento, M.: Thirty years of graph matching in pattern recognition. International journal of pattern recognition and artificial intelligence **18**(03), 265–298 (2004)

[5] Cortés, X., Conte, D., Cardot, H., Serratosa, F.: A deep neural network architecture to estimate node assignment costs for the graph edit distance. In: Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR). pp. 326–336. Springer (2018)

[6] Cortés, X., Serratosa, F.: Learning graph-matching edit-costs based on the optimality of the oracle's node correspondences. Pattern Recognition Letters **56**, 22–29 (2015)

[7] Cortés, X., Serratosa, F.: Learning graph matching substitution weights based on the ground truth node correspondence. International Journal of Pattern Recognition and Artificial Intelligence **30**(02), 1650005 (2016)

[8] Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence (2), 224–227 (1979)

[9] Dunn, J.C.: Well-separated clusters and optimal fuzzy partitions. Journal of Cybernetics **4**(1), 95–104 (1974). https://doi.org/10.1080/01969727408546059, https://doi.org/10.1080/01969727408546059

[10] Fowlkes, E.B., Mallows, C.L.: A method for comparing two hierarchical clusterings. Journal of the American statistical association **78**(383), 553–569 (1983)

[11] Gao, X., Xiao, B., Tao, D., Li, X.: A survey of graph edit distance. Pattern Anal. Appl. **13**(1), 113–129 (Feb 2010). https://doi.org/10.1007/s10044-008-0141-y, http://dx.doi.org/10.1007/s10044-008-0141-y

[12] Goodman, L.A., Kruskal, W.H.: Measures of association for cross classifications. In: Measures of association for cross classifications, pp. 2–34. Springer (1979)

[13] Hubert, L., Schultz, J.: Quadratic assignment as a general data analysis strategy. British journal of mathematical and statistical psychology **29**(2), 190–241 (1976)

[14] Jain, A.K., Dubes, R.C.: Algorithms for clustering data (1988)

[15] Leordeanu, M., Sukthankar, R., Hebert, M.: Unsupervised learning for graph matching. International journal of computer vision **96**(1), 28–45 (2012)

[16] Lloyd, S.: Least squares quantization in pcm. IEEE Trans. Inf. Theor. **28**(2), 129–137 (Sep 2006). https://doi.org/10.1109/TIT.1982.1056489, `http://dx.doi.org/10.1109/TIT.1982.1056489`

[17] Martineau, M., Raveaux, R., Conte, D., Venturini, G.: Learning error-correcting graph matching with a multiclass neural network. Pattern Recognition Letters (2018)

[18] Moreno-Garca, Carlos Francisco, C.X., Serratosa, F.: A graph repository for learning error-tolerant graph matching. Syntactic and Structural Pattern Recognition (2016)

[19] Neuhaus, M., Bunke, H.: Self-organizing maps for learning the edit costs in graph matching. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **35**(3), 503–514 (2005)

[20] Neuhaus, M., Bunke, H.: Automatic learning of cost functions for graph edit distance. Information Sciences **177**(1), 239–247 (2007)

[21] Rand, W.M.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association **66**(336), 846–850 (1971)

[22] Raveaux, R., Martineau, M., Conte, D., Venturini, G.: Learning graph matching with a graph-based perceptron in a classification context. In: International Workshop on Graph-Based Representations in Pattern Recognition. pp. 49–58. Springer (2017)

[23] Sanfeliu, A., Alquézar, R., Andrade, J., Climent, J., Serratosa, F., Vergés-Llahí, J.: Graph-based representations and techniques for image processing and image analysis. Pattern Recognition **35**, 639–650 (2002)

[24] Santacruz, P., Serratosa, F.: Learning the sub-optimal graph edit distance edit costs based on an embedded model. In: S+SSPR. Lecture Notes in Computer Science, vol. 11004, pp. 282–292. Springer (2018)

[25] Serratosa, F.: Fast computation of bipartite graph matching. Pattern Recognition Letters **45**, 244–250 (2014)

[26] Serratosa, F.: Speeding up fast bipartite graph matching through a new cost matrix. International Journal of Pattern Recognition and Artificial Intelligence **29**, 1550010 (11 2014). https://doi.org/10.1142/S021800141550010X

[27] Serratosa, F.: A methodology to generate attributed graphs with a bounded graph edit distance for graph-matching testing. IJPRAI **32**(11), 1850038 (2018)

[28] Serratosa, F.: Graph edit distance: restrictions to be a metric. Pattern Recognition **90**, 250–256 (2019)

[29] Serratosa, F., Cortés, X.: Interactive graph-matching using active query strategies. Pattern Recognition **48**(4), 1364–1373 (2015)

[30] Stauffer, M., Tschachtli, T., Fischer, A., Riesen, K.: A survey on applications of bipartite graph edit distance. In: GbRPR. Lecture Notes in Computer Science, vol. 10310, pp. 242–252 (2017)

[31] Vento, M.: A one hour trip in the world of graphs, looking at the papers of the last ten years. In: Kropatsch, W.G., Artner, N.M., Haxhimusa, Y., Jiang, X. (eds.) Graph-Based Representations in Pattern Recognition. pp. 1–10. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)

Predicting mortality and duration of mechanical ventilation of Acute Respiratory Distress Syndrome patients: a supervised learning approach

Mohammed Sayed and David Riano ⋆

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
mgamal.sayed@urv.cat,david.riano@urv.cat

Abstract. The acute respiratory distress syndrome (ARDS) is a frequent type of respiratory failure observed in intensive care units (ICU). ARDS is a clinically and biologically heterogeneous disorder associated with many disease interactions that injure the lung, culminating in increased non-hydrostatic extravascular lung water, reduced compliance, and severe hypoxemia. In spite of the enhanced understanding of molecular mechanisms, advances in ventilation strategies, and general care of the critically ill patient, mortality ratio remains unacceptably high. The difficulty in early ARDS stage originates from its complicated and heterogeneous nature. The current Berlin classification proposes three severity levels of ARDS (mild, moderate, and severe). In this study, we integrate knowledge of the heterogeneity of ARDS patients into predictive models using Light Gradient Boosting Machine (LightGBM) and Random Forest (RF) algorithms. The prediction of mechanical ventilation duration and mortality are two unsolved issues in ARDS and they are essential attributes for correcting the level of care and ICU management. LightGBM and RF algorithms were used to predict the duration of mechanical ventilation and mortality within each one of the Berlin severity groups with the database MIMIC III (MetaVision). Our results show that lightGBM is more powerful than RF in predicting both of the duration of mechanical ventilation and mortality within each one of the Berlin severity groups in ARDS patients. Also, the prediction performances of LightGBM and RF were compared and analyzed with respect to each Berlin severity group in ARDS patients.

Keywords: Knowledge extraction from health care databases and medical records, Light Gradient Boosting Machine, Random Forest, Prediction, Acute Respiratory Distress Syndrome.

⋆ PhD advisor: David Riano

1 Introduction

The acute respiratory distress syndrome (ARDS) is an essential cause of morbidity and mortality in the USA and worldwide [1]. ARDS is a life-threatening form of respiratory failure characterized by inflammatory pulmonary edema leading to severe hypoxemia [2]. Its overall prevalence in the intensive care unit (ICU) was recently reported by 10.4% [1]. ARDS was first described in 1967 [3].

The Berlin definition of ARDS was declared in 2012 and it replaced the previous one of the American-European Consensus Conference (AECC) that was in 1994. It identifies three mutually exclusive categories of severity which are defined for patients with a positive end expiratory pressure (PEEP) of 5 cmH20 or greater: mild, with PaO2/FiO2 in (200-300) mmHg; moderate, with PaO2/FiO2 in (100-200) mmHg; and severe with PaO2/FiO2 less than or equal 100 mmHg [4,5]. Patients in these severity classifications show different mortality ratios [6]: 27% for mild (24- 30%, 95% CI), 32% for moderate (29-34%), and 45% for severe (42-48%), with a P-value ¡ 0.001. Besides, mortality prediction with respect to Berlin definition severity groups outperformed previous ARDS classifications with an AUC (ROC) of 0.577 (0.561-0.593, 95% CI) [6]. In spite of that, Lorenzo Del Sorbo et al. [7] reported that several important issues were addressed in the Berlin definition of ARDS. Also, a recent publication [8] addresses the risks of the Berlin definition. Precisely, it discusses that mild ARDS category may be severe in terms of level of care and outcome. These two important health care quality attributes (i.e., level of care and outcome) can be measured in terms of the duration of mechanical ventilation in the ICU, and the mortality rate, among other clinical parameters.

MIMIC-III is a large, publicly-available database including de-identified health-related data of approximately sixty thousand admissions of patients who stayed in ICUs of the Beth Israel Deaconess Medical Center in Boston [9]. In MIMIC-III, original data (years 2001-2008) were extended with the MetaVision data (years 2008-2012). MetaVision data contains data of 23,024 ARDS patients, as described according to Berlin definition.

Machine learning could be used to predict the unknown data [10]. The basic process is to extract the feature of the training data, design the classifier and obtain the prediction model by supervised learning [11,12]. In order to develop the prediction models for ARDS patients, two machine learning algorithms, LightGBM and RF, with a python wrapper, were run to predict the duration of mechanical ventilation and mortality within each one of the Berlin severity groups of ARDS patients, but also for all the patients regardless of their severity level. Their prediction performances were compared and analyzed within each one of the Berlin severity groups.

2 Methods

LightGBM [13] is an open source Gradient Boosting Decision Tree (GBDT) algorithm by Microsoft. It uses the histogram-based algorithm [14,15] to speed up the training process and reduce memory consumption. It combines advanced network communication to optimize parallel learning, called parallel voting decision tree algorithm. Dividing the training data into multiple machines, the local voting decision to select the top-k attributes and the global voting decision to receive the top-2k attributes in each iteration are performed. LightGBM uses the leaf-wise strategy to find a leaf with largest splitter gain [16].

Random Forests are a scheme proposed by Leo Breiman in the 2000s for building a predictor ensemble with a set of decision trees that grow in randomly selected subspaces of data. In spite of increasing interest and practical use, there has been little exploration of the statistical properties of random forests, and also little is known about the mathematical forces driving the algorithm [17].

In order to predict the duration of mechanical ventilation and the mortality of ARDS patients in ICUs, the clinical features described in [18,19,20] were selected. These are: age (years), PEEP (cmH20), PaO2/FiO2, Mean heart rate (beats per hour), Mean respiratory rate (breaths per hour), and Number of ventilation actions. Apart of these, for each patient, we counted with the additional 3 features duration of mechanical ventilation (hours), survival (0/1), and the Berlin definition classification (mild, moderate, or severe).

In this study, 70% of the eligible patients were randomly chosen as the training set, which was used to train the machine learning models. The rest 30%of the patients were utilized as the holdout test set. Various models were then built on the training set and the hold-out test set was used to evaluate the models performance by four metrics. Accuracy, F1-score, and area under the ROC curve (AUC) were used to evaluate each models quality in predicting survival (qualitative). The mean absolute error (MAE) was used to evaluate each models quality in predicting the duration of mechanical ventilation (quantitative).

3 Results

3.1 Predicting the mortality of ARDS patients within each one of the Berlin severity groups

Predictive models were built for all the cases together and for cases in each severity class, separately. Performances of these predictive models are shown in Tables 1 (LightGBM model) and 2 (RF model).

For the LightGBM model, we can observe that the global model on all the ARDS cases has better quality results than when the model is trained

separately for each one of the Berlin severity levels. In general, the moderate model shows a better performance than the models obtained for mild and severe ARDS. When all the cases are considered without a distinction of their severity, accuracy, F1-score, and AUC values reach a close to the exact prediction (100%), as the last column of Table1 shows.

Table 1. Prediction performance results of LightGBM in terms of Accuracy, F1-score, and area under the ROC curve (AUC)

LightGBM	MILD	MODERATE	SEVERE	ALL
ACC	0.952	0.968	0.953	0.990
FS	0.889	0.940	0.921	0.981
AUC	0.969	0.987	0.984	0.999

Table 2. Prediction performance results of RF in terms of Accuracy, F1-score, and area under the ROC curve (AUC)

RF	MILD	MODERATE	SEVERE	ALL
ACC	0.925	0.962	0.961	0.925
FS	0.826	0.927	0.937	0.826
AUC	0.856	0.937	0.943	0.856

For the RF model in Table 2, when all the cases are used in the training, the predictive quality is equivalent to mild cases, but significantly better for moderate and severe ARDS. This describes an inversion of the concepts of ARDS severity and ARDS prediction in the sense that the higher the severity of ARDS is, the best predictive models of evolution (i.e., mortality) are obtained. In other words, predicting the mortality for mild ARDS cases is harder than for moderate ARDS cases, and this, in its turn, is more difficult than for severe ARDS.

Crossed comparisons between LightGBM and RF models for patients each severity class and for all cases indicate that significant improvements in AUC over existing predictive models are obtained when using LightGBM rather than when we use RF, except for the accuracy and F1-score of severe cases, where Rf seems to perform slightly better (1%).

3.2 Predicting the duration of ventilation of ARDS patients within each one of the Berlin severity groups

Mechanical ventilation of an ARDS patient is important, and the duration of this ventilation is essential not only for the care of the patient, but also for the

correct management of ICUs. When we used the MIMIC III data to predict the number of hours of ventilation of ARDS patients, several predictive models were built for all cases and for cases in each Berlin severity class, separately. The performances of the predictive models evaluated are listed in Tables 3 and 4. Again, Table 3 describes the results of the LightGBM models, and Table 4 the results of the RF models, in terms of the mean absolute error (MAE) in hours. So, the mean absolute error of predicting the duration of ventilation of mild ARDS patients is 45 hours (almost 2 days) if we use the LightGBM method, and close to 30 hours (a bit more than one day) if we consider moderate or severe patients, but only a 15-hour error if we do not make a distinction of the severity of the cases.

Comparisons between models for cases in each severity class and for all cases together indicate that the prediction errors of the models obtained with LightGBM are lower than the errors produced by the RF models. However, the overall behavior of both LightGBM models and RF models are equivalent since, their lower errors are observed when all the patients are considered together, then for moderate ARDS cases, then for severe cases, and finally for mild cases. In other words, predicting the required number of hours of mechanical ventilation is more error prone for mild ARDS cases, than for the rest.

Table 3. Prediction performance results of LightGBM in terms of MAE

LightGBM	MILD	MODERATE	SEVERE	ALL
MAE	45.029	28.710	30.490	15.723

Table 4. Prediction performance results of RF in terms of MAE

RF	MILD	MODERATE	SEVERE	ALL
MAE	49.93	41.61	43.36	28.27

4 Discussion

Using available clinical parameters, our analysis considered three different classes of ARDS, according to Berlin definition. Our study applied Light-GBM and RF to all ARDS patients but also to patients in each one of the three Berlin severity classes, and built separate predictive models for each group regarding to mortality and duration of ventilation. Our results indicate that significant improved performances of prediction can be obtained for all ARDS patients rather than for the three Berlin severity classes in isolation (especially, for the mild class) using LightGBM which is more successful than

RF. Two possible interpretations are possible: on the one hand, the heterogeneous nature of ARDS patients within each severity class, with respect to mortality and ventilation time patient needs (i.e., the Berlin classification is not appropriate to consider mortality and duration of mechanical ventilation). On the other hand, the used data splitting into training set (70%) and test set (30%). Since only one random split was considered, this could cause a bias in the results. Further refinements of the method may be conducted to completely understand some particular results.

Several prediction scores have been developed to evaluate ARDS prognosis and risk of death, such as the Lung Injury Score (LIS) [21], Lung Injury Prediction Score (LIPS) [22], APPS (age, plateau, PaO2/FiO2) score [20], Early Acute Lung Injury (EALI) [23], and Modified ARDS Prediction Score (MAPS) [24]. However, the predictive validities of these ARDS scoring tools have been shown to be moderate. For instance, LIPS discriminated patients who developed Acute Lung Injury (ALI) from those who did not with an AUC of 0.80 (95% CI, 0.770.84). MAPS was shown to have a similar AUC of 0.79 (95% CI, 0.72 - 0.87) in predicting ARDS development, and the reported EALI AUC was 0.85 (95% CI: 0.80-0.91), on the training set, for identifying patients who progressed to acute lung injury [23].

With the appearance of the Berlin definition, the predictive validity of LIS was found to be limited, with an AUC of 0.60 (95% CI 0.55 to 0.65) in identifying mortality [25]. Similarly, APPS was found to have an AUC of 0.8 for predicting ARDS mortality [20]. General illness severity scores, such as SAPS, SAPS II, APACHE II/III, and MPM, were also examined in their ability in helping to recognize ARDS early, but only similar moderate performances have been reported [20,23,24]. Improved predictive validity is required to enable reliable early identification and management of patients at risk for ARDS. We assume that an obstacle to improved predictive performance of existing scoring tools may be the heterogeneity of the ARDS populations used to derive these models. Studies have indicated that ARDS is a highly heterogeneous syndrome [26,27,28]. Such heterogeneity in population leads to heterogeneity in relationships between explanatory and response variables within partitions, resulting in serious challenges in predictive model building [29].

The competency of our study is the integration of knowledge of the heterogeneity of ARDS patients into predictive models using the LightGBM and the RF algorithms. The obtained results also showed a high competency of these models in the performance of prediction, rather than previous studies such as [4,19].

5 Conclusions

In this study, we integrated knowledge of the heterogeneity of ARDS patients into predictive model building using LightGBM and RF algorithms. Our re-

sults showed clearly that lightGBM is more powerful than RF in predicting both of the duration of mechanical ventilation and mortality within each one of the Berlin severity groups in ARDS patients. In spite of that, predictive validity of LightGBM showed notably inferior prediction performance for the three Berlin class-severity (especially, the mild class), if it is compared with all ARDS patients as a whole group. Such results encourage us to expand our study to improve the prediction performance of ARDS using more clinical parameters. Our future work will include efforts on exploring other clinical parameters and using cross validation in order to enhance the performance of prediction models.

References

[1] G. Bellani, J. G. Laffey, T. Pham, E. Fan, L. Brochard, A. Esteban, L. Gattinoni, F. Van Haren, A. Larsson, D. F. McAuley, *et al.*, "Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries," *Jama*, vol. 315, no. 8, pp. 788–800, 2016.

[2] L. B. Ware and M. A. Matthay, "The acute respiratory distress syndrome," *New England Journal of Medicine*, vol. 342, no. 18, pp. 1334–1349, 2000.

[3] D. Ashbaugh, D. Bigelow, T. Petty, and B. Levine, "Ashbaugh dg, bigelow db, petty tl, levine be. acute respiratory distress in adults. the lancet, saturday 12 august 1967.," *Critical care and resuscitation: journal of the Australasian Academy of Critical Care Medicine*, vol. 7, no. 1, pp. 60–61, 2005.

[4] A. D. T. Force, V. Ranieri, G. Rubenfeld, *et al.*, "Acute respiratory distress syndrome," *Jama*, vol. 307, no. 23, pp. 2526–2533, 2012.

[5] N. D. Ferguson, E. Fan, L. Camporota, M. Antonelli, A. Anzueto, R. Beale, L. Brochard, R. Brower, A. Esteban, L. Gattinoni, *et al.*, "The berlin definition of ards: an expanded rationale, justification, and supplementary material," *Intensive care medicine*, vol. 38, no. 10, pp. 1573–1582, 2012.

[6] A. Definition Task Force, V. Ranieri, G. Rubenfeld, B. Thompson, N. Ferguson, E. Caldwell, E. Fan, L. Camporota, and A. Slutsky, "Acute respiratory distress syndrome: the berlin definition," *JaMa*, vol. 307, no. 23, pp. 2526–2533, 2012.

[7] L. Del Sorbo, V. M. Ranieri, and N. D. Ferguson, "The berlin definition met our needs: yes," 2016.

[8] R. Pirracchio and M. A. Gropper, "Heterogeneity in intensive carelow severity does not mean low risk!," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 130, no. 2, pp. 190–191, 2019.

[9] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, p. 160035, 2016.

[10] E. Alpaydin, *Introduction to machine learning.* MIT press, 2014.

[11] D. Oliva and E. Cuevas, *Advances and applications of optimised algorithms in image processing.* Springer, 2017.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[13] Q. Meng, G. Ke, T. Wang, W. Chen, Q. Ye, Z.-M. Ma, and T.-Y. Liu, "A communication-efficient parallel algorithm for decision tree," in *Advances in Neural Information Processing Systems*, pp. 1279–1287, 2016.

[14] S. Ranka and V. Singh, "Clouds: A decision tree classifier for large datasets," in *Proceedings of the 4th Knowledge Discovery and Data Mining Conference*, vol. 2, p. 8, 1998.

[15] R. Jin and G. Agrawal, "Communication and memory efficient parallel decision tree construction," in *Proceedings of the 2003 SIAM International Conference on Data Mining*, pp. 119–129, SIAM, 2003.

[16] D. Wang, Y. Zhang, and Y. Zhao, "Lightgbm: an effective mirna classification method in breast cancer patients," in *Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics*, pp. 7–11, ACM, 2017.

[17] G. Biau, "Analysis of a random forests model," *Journal of Machine Learning Research*, vol. 13, no. Apr, pp. 1063–1095, 2012.

[18] E. Fan, D. Brodie, and A. S. Slutsky, "Acute respiratory distress syndrome: advances in diagnosis and treatment," *Jama*, vol. 319, no. 7, pp. 698–710, 2018.

[19] T. Aline, M.-C. Farah, A. Hassan, C. Aly, F. Ziad, and D. Jacques, "Data fusion for predicting ards using the mimic ii physiological database," in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1–5, IEEE, 2016.

[20] J. Villar, A. Ambrós, J. A. Soler, D. Martínez, C. Ferrando, R. Solano, F. Mosteiro, J. Blanco, C. Martín-Rodríguez, M. d. M. Fernández, *et al.*, "Age, pao2/fio2, and plateau pressure score: a proposal for a simple outcome score in patients with the acute respiratory distress syndrome," *Critical care medicine*, vol. 44, no. 7, pp. 1361–1369, 2016.

[21] J. F. Murray, M. A. Matthay, J. M. Luce, M. R. Flick, *et al.*, "An expanded definition of the adult respiratory distress syndrome," *Am Rev Respir Dis*, vol. 138, no. 3, pp. 720–723, 1988.

[22] O. Gajic, O. Dabbagh, P. K. Park, A. Adesanya, S. Y. Chang, P. Hou, H. Anderson III, J. J. Hoth, M. E. Mikkelsen, N. T. Gentile, *et al.*, "Early identification of patients at risk of acute lung injury: evaluation of lung injury prediction score in a multicenter cohort study," *American journal of respiratory and critical care medicine*, vol. 183, no. 4, pp. 462–470, 2011.

[23] J. E. Levitt, C. S. Calfee, B. A. Goldstein, R. Vojnik, and M. A. Matthay, "Early acute lung injury: criteria for identifying lung injury prior to the need for positive pressure ventilation," *Critical care medicine*, vol. 41, no. 8, p. 1929, 2013.

[24] J. Xie, L. Liu, Y. Yang, W. Yu, M. Li, K. Yu, R. Zheng, J. Yan, X. Wang, G. Cai, *et al.*, "A modified acute respiratory distress syndrome prediction score: a multicenter cohort study in china," *Journal of thoracic disease*, vol. 10, no. 10, p. 5764, 2018.

[25] K. N. Kangelaris, C. S. Calfee, A. K. May, H. Zhuo, M. A. Matthay, and L. B. Ware, "Is there still a role for the lung injury score in the era of the berlin definition ards?," *Annals of intensive care*, vol. 4, no. 1, p. 4, 2014.

[26] C. S. Calfee, K. Delucchi, P. E. Parsons, B. T. Thompson, L. B. Ware, M. A. Matthay, N. A. Network, *et al.*, "Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials," *The Lancet Respiratory Medicine*, vol. 2, no. 8, pp. 611–620, 2014.

[27] P. Sinha, K. L. Delucchi, B. T. Thompson, D. F. McAuley, M. A. Matthay, C. S. Calfee, N. A. Network, *et al.*, "Latent class analysis of ards subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (sails) study," *Intensive care medicine*, vol. 44, no. 11, pp. 1859–1869, 2018.

[28] Z. Zhang, "Identification of three classes of acute respiratory distress syndrome using latent class analysis," *PeerJ*, vol. 6, p. e4592, 2018.

[29] A. Karpatne, A. Khandelwal, S. Boriah, and V. Kumar, "Predictive learning in the presence of heterogeneity and limited training data," in *Proceedings of the 2014 SIAM International Conference on Data Mining*, pp. 253–261, SIAM, 2014.

Privacy-preserving Mechanisms in Smart Metering

Bastian Stölb [*]

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Catalonia
`bastiannikolas.stolb@estudiants.urv.cat`

1 Introduction

By the end of 2018 roughly 44 percent of European households were equipped with a Smart Meter (SM). According to Ryberg from Berg Insight[11] reports, by the year 2023 the distribution will be around 71 percent, which means that over 200 million homes in Europe will be equipped with smart meters.

With SMs, the power usage can be observed in real-time by the energy providers, thus allowing to produce as much electricity as necessary in a specific time. However, this increase of efficiency comes at the costs of privacy. Specifically, the number of people in an accommodation at a given moment can be found out by using Smart Meter data. This invites burglars when the power consumption of a flat is very little to no consumption at all.

Also, the habits of the residents can lead to serious privacy concerns: as stated by Garcia and Jacobs in [9] a person who wakes up at five o'clock in the morning in combination with a foreign name could possibly be identified as a religious Muslim.

At a lower level, a so-called Nonintrusive Appliance Load Monitoring (NALM) [10] can even observe the electricity usage in detail. By using signatures of electronic devices (iron, refrigerator, dish washer, microwave, hair dryer, etc.) it can be determined which device is turned on or off, with notable precision.

Other privacy risks and numerous conflicts of interests are outlined by Anderson and Fuloria in [2]. Among others, the researchers give the following suggestions: "smart meter data should belong to the customer[2]". The electricity provider needs access to these data only for supply and accounting reasons. A common database is refused by the authors; instead they demand "a framework of standards that allow data to be shared between energy suppliers, distributors and management companies[3]".

---

[*] PhD advisors: Josep Domingo-Ferrer, David Sánchez and Jordi Soria-Comas
[2] [2], p. 16
[3] [2], p. 16

2 Non-Cryptographic Smart Meter Privacy

The communication of smart meters with utility companies takes place over a public network (the Internet) and is regarded as unsafe unless appropriate countermeasures are taken. That is why a lot of work in this area has been performed and is still ongoing on protecting privacy. Most researchers focus on cryptographic solutions, mainly Partially Homomorphic Encryption (PHE) and Fully Homomorphic Encryption (FHE). Although cryptography-based methods are appealing, one should consider the limited computation capacity of a smart meter. With this in mind, non-cryptographic solutions seem worth exploring.

2.1 Privacy models

Non-cryptographic solutions are normally oriented to satisfying a privacy model, that is, a privacy condition. As stated by Domingo-Ferrer and Soria-Comas in [6], there are four major privacy models:
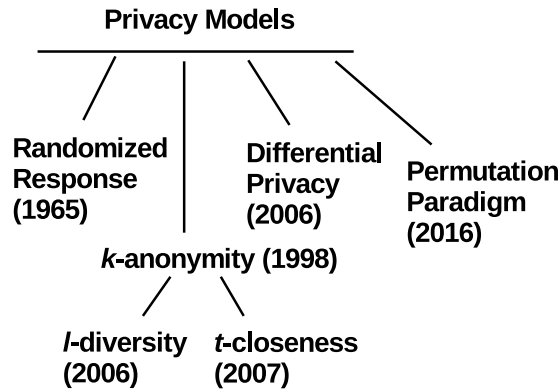

Fig. 1. Privacy Models

- *Randomized Response.* Randomized Response (RR) was invented by Warner in 1965 [16]. Scharrer describes the technique in [12] as follows: "The idea of this method is to ask sensitive questions (e.g. are you pregnant?), while assuring the privacy of the respondent. Therefore the survey is combined with a random experiment: the respondent has to answer one of three questions, randomly selected and not presented to the interviewer. Thus the interviewer only knows the answer, without the corresponding question. By knowing the probability of the questions, the true answers to the sensitive question can be estimated."

- *k-anonymity.* k-anonymity [14] was designed for database applications. A released data set is said to be k-anonymous if any combination of values of quasi-identifier attributes is shared by at least k records. This provides protection against re-identification of the subjects to whom the records correspond. Bohli et al. [3] stated, that "The concept of k-anonymity is not suited to the smart metering problem, as there is no central entity releasing the smart meter readings [...]".
- *Differential Privacy.* According to [7] and [6], a query to a database is said to be differentially private if from its result it is not possible to notice the presence or absence of any specific record in the database. Usually the privacy is protected by adding noise to the real query result.
- *Permutation Paradigm.* In [5], the authors presented the permutation paradigm, whereby they showed that any statistical disclosure control methods essentially consists of permutation plus perhaps a small noise addition. [6] describes the two steps permutation and noise addition as follows: "Each attribute X of the original dataset is permuted into the corresponding Z. Thus, the data set X is transformed into a data set Z. [...] Noise is added to each value of Z to obtain the anonymized data set Y [...]".

2.2 Data aggregation

Aggregation is a common principle to achieve privacy models. Aggregation is also useful to reduce costs for computation and communication. Both are limiting factors as smart meters have to deal with constrained resources. A nice overview about aggregation is given by Erkin and Tsudik in [8]:

- *Spatial aggregation.* In this approach smart meters are (geographically) clustered. The data of an urban district (e.g.) can be accumulated and presented to the energy supplier. This guarantees that individual households are protected, while load-balancing is still possible at the same time. Self-supplying, outlandish locations can benefit from monitoring the energy consumption. For example a person can check if there is enough energy available before plugging-in another electrical device.
- *Temporal aggregation.* Temporal aggregation means that smart meters withhold data until a specific time interval has passed. Fine granular data is summed up by the smart meter itself and protects the privacy of the customers. For the purpose of accounting, monthly aggregation is sufficient. However, many energy providers would like to have a consumption report at least on a daily basis.
- *Spatio-temporal aggregation.* In this hybrid setting, both approaches are combined. The data of a single SM is spatially aggregated at a specific time for operational intentions. For the aim of billing the measurements of an individual SM are aggregated over time.

3 Related Work

According to Bohli, Sorge and Ugus [3], privacy can be defined by using a "right-or-left type of [cryptographic] game[4]". Two proposals are offered: one of them includes a Trusted Third Party (TTP), while the other does not. In the first scenario the researchers suggest to use a TTP for aggregation purposes. The data are encrypted by the smart meter before it sends them to the TTP. The second outline was designed to be more frugal: every SM perturbates the measured values by adding random noise (with a specific distribution). In case of malfunction of a single SM, the energy provider is aware of missing data and the total usage can be roughly estimated. To achieve sufficient privacy, it must be assured that the added random values are considerably large, which is a drawback.

In [1] the authors suggest the usage of a differential privacy model. The proposal is straightforward, economic and easy to implement. There is no need for a Trusted Third Party. Individual smart meters add Laplacian noise to the measured data, before a stream cipher is applied. This results in quite low computational costs. The smart meters are clustered[5] and aggregated. The aggregator receives the accumulated values of all smart meters in the cluster. For the aggregator it is impossible to learn the consumption of an individual smart meter (at a specific time). The coarse resolution is disadvantageous. Also this approach does not cope with malfunctions. In case of a single smart meter fails, the data of the whole cluster are lost, because of the stream cipher. Assuming the faulty smart meter is able to store the value and retransmit it later, the reliability of the proposed scheme can be improved.

A recent paper is [13]. The author presents an overview about ongoing research in the area of SM. To that end, he analyzed 53 papers (thereof 8 surveys) published in the last 9 years. He divides the studies into two main groups: attributable (with and without aggregation) or non-attributable (for accounting and operational purposes). He investigates whether a Trusted Third Party (TTP) is used. New approaches, like incentive-based or rewarding schemes, where users voluntarily share their data with their utility provider are emphasized. Although the paper provides a great overview about the state-of-the-art in smart meter research, it does not distinguish between cryptographic and non-cryptographic solutions. Then we have found the following two papers using randomized response.

Wang et al. described a model [15], in which a single SM can send the true data with a known probability. This approach requires so called Load Serving Entities (LSE). These LSEs can calculate the aggregated usage of a region, by using a statistical inference algorithm.

---

[4] [3], p. 2
[5] An urban district consists of several hundreds or thousands of smart meters.

A recent publication from 2018 by Cao et al. [4] suggests to add RR noise into the signatures matrix of the behavior of the electronic devices. Furthermore the researchers recommend a technique to model the signature of behavior, by using sparse coding. Thus a dictionary is generated after a short training period, containing the attributes of the electronic devices.

4 Our Proposed Scheme

4.1 Goals

We will propose a non-cryptographic solution that satisfies the following goals:

- privacy
- simplicity (lightweight, no need for Trusted Third Party)
- support individual readings (without aggregation)
- low computation overhead
- low communication overhead
- integrity
- high accuracy

The approach will be demonstrated with real data (ESSnet Big Data[6]) and using realistic Smart Meter architecture (with respect to scalability).

4.2 Threats

The following threats ought to be considered:

- eavesdropping
- tampering
- internal attacker

However, the smart meters themselves will be assumed to be trusted devices.

4.3 Rationale

The following diagram compares cryptographic and non-cryptographic solutions. Our aim is to develop a scheme by using RR, which can handle individual readings and maintain privacy at the same time.

---

[6] https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WP3_Report_1_1

CRYPTO — individual readings with (homomorphic) aggregation →

NON-CRYPTO — time-aggregated readings (noise is added to them) → collect and protect non-aggregated individual readings (with RR?) →
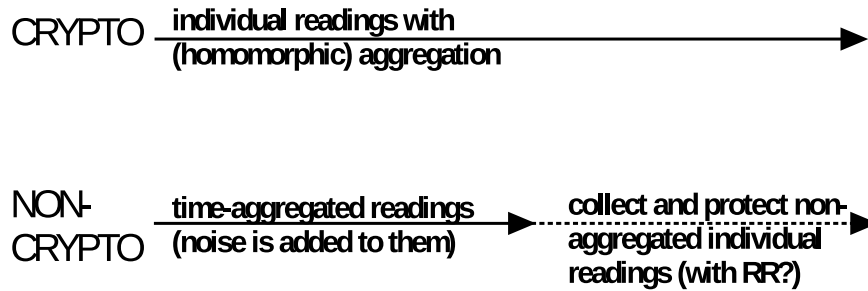
Fig. 2. Cryptographic vs. Non-Cryptographic solutions

4.4 Preliminary ideas

The proposed scheme by Wang et al. [15], relies on the Gaussian Mixture Model (GMM), assuming that "[...] the readings of smart meters are i.i.d.[7], and their distributions are modeled with a normal distribution parameterized by mean and variance."[8]. In this approach real SM data is mixed with "[...] faked readings from $K - 1$ pre-determined distributions."[9] As the authors stated, in a general, more complex setting the data of individual SMs can be mixed with different uncertainties. This results in an improved privacy preservation. The GMM is often applied in the area of Machine Learning (ML). Therefore connections will be investigated between RR and ML with respect to privacy preservation in smart metering. We will take as starting point the approach by Wang, due to its simplicity.

References

[1] G. Ács and C. Castelluccia. I have a dream! (differentially private smart metering). In T. Filler, T. Pevný, S. Craver, and A. Ker, editors, *Information Hiding*, pages 118–132, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[2] R. Anderson and S. Fuloria. On the security economics of electricity metering. In *On the security economics of electricity metering*, 2010.

[3] J. Bohli, C. Sorge, and O. Ugus. A privacy model for smart metering. In *2010 IEEE International Conference on Communications Workshops*, pages 1–5, May 2010.

[4] H. Cao, S. Liu, Z. Guan, L. Wu, H. Deng, and X. Du. An efficient privacy-preserving algorithm based on randomized response in iot-based smart grid. In *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big*

---

[7] independent and identically distributed
[8] [15], p. 1318
[9] [15], p. 1318

Data Computing, Internet of People and Smart City Innovation (Smart-World/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), pages 881–886, Oct 2018.

[5] J. Domingo-Ferrer and K. Muralidhar. New directions in anonymization: Permutation paradigm, verifiability by subjects and intruders, transparency to users. *Information Sciences*, 337-338:11 – 24, 2016.

[6] J. Domingo-Ferrer and J. Soria-Comas. Connecting randomized response, post-randomization, differential privacy and t-closeness via deniability and permutation. *CoRR*, abs/1803.02139, 2018.

[7] C. Dwork. Differential privacy. In M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, editors, *Automata, Languages and Programming*, pages 1–12, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[8] Z. Erkin and G. Tsudik. Private computation of spatial and temporal power consumption with smart meters. In F. Bao, P. Samarati, and J. Zhou, editors, *Applied Cryptography and Network Security*, pages 561–577, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[9] F. D. Garcia and B. Jacobs. Privacy-friendly energy-metering via homomorphic encryption. In J. Cuellar, J. Lopez, G. Barthe, and A. Pretschner, editors, *Security and Trust Management*, pages 226–238, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[10] G. W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, Dec 1992.

[11] T. Ryberg. Smart metering in europe, Nov. 2018. Berg Insight at http://www.berginsight.com/ReportPDF/ProductSheet/bi-sm14-ps.pdf, accessed 2019-04-22.

[12] N. Scharrer. Die entwicklung der randomized response technik bis hin zur zusammenhangsanalyse, 2014. Bachlor Thesis, https://epub.ub.uni-muenchen.de/21449/1/BA_Scharrer.pdf, accessed 2019-04-25.

[13] S. Sultan. Privacy-preserving metering in smart grid for billing, operational metering, and incentive-based schemes: A survey. *Computers & Security*, 84:148 – 165, 2019.

[14] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[15] S. Wang, L. Cui, J. Que, D. Choi, X. Jiang, S. Cheng, and L. Xie. A randomized response model for privacy preserving smart metering. *IEEE Transactions on Smart Grid*, 3(3):1317–1324, Sep. 2012.

[16] S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965. PMID: 12261830.

Decentralized access control system for Low Emission Zones

Carles Anglès-Tafalla [*]

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
`carles.angles@urv.cat`

The high levels of pollution, mostly to vehicular traffic congestion, have become a serious problem in large cities around the globe. In the view of their effectiveness, Governments in Europe have started to adopt Low Emission Zones, i.e. areas where restrictions or surcharges are accordingly applied to vehicles' emissions, as a way to put a brake on this problematic. However, the intrusive nature of the currently deployed automated systems has given rise to public concern for their users' privacy. Although several approaches reduced the use of cameras from full vehicle tracking to only identifying fraudulent users, current works still pose a strong dependence on centralized entities in charge of acknowledging vehicles' access data and determining their fees. According to these centralization issues, we propose an efficient privacy-preserving solution for controlling the access to LEZ*s*, whose fundamental principle is the use smart contracts to omit third parties from participating in payment related processes in favor of the decentralized validation system the blockchain paradigm poses.

1 Introduction

The registered high levels of environmental pollution, due in large part to urban traffic congestion, have become a serious problem for large cities all around the world. In the downtown of these metropolitan areas, pollution levels far exceed some of the limits established by the World Health Organization [1], posing a danger to the health of their citizens. In order to address this problem, government administrations have begun to implement measures to encourage the rational use of vehicles, which include, among others, restrictions on polluting vehicles' driving, special lanes for High Occupancy Vehicles (HOV) or definition of Low Emission Zones (LEZ).

Among these proposals, the implementation of LEZs, which consists of a delimited area where some restrictions are applied to drivers in accordance

---

[*] PhD advisor: Jordi Castellà Roca and Alexandre Viejo

with their vehicles' emissions, is one of the measures that has proliferated the most. Sweden, Italy, the Netherlands, the United Kingdom or Germany are clear examples of countries implementing these kind of schemes in their cities.

In the view of that trend, the need to implement access control systems for LEZ which enables compliance with such vehicular restrictions and toll payments arouse. In that matter, the most cited case in the literature is the London LEZ and its controversial control system [2]. The backbone of such system is composed by a network of more than 300 cameras distributed around the city center, whose purpose is to indiscriminately photograph the license plates of all the vehicles circulating inside it. Later, the vehicles are identified in order to verify if they are paying their corresponding fees. A similar system has been running in Stockholm since 2007 in order to control vehicle accesses into the LEZ which covers the entire city center. As in London, its access control system uses automatic number plate recognition based on cameras which automatically register the vehicles circulating at all entrance points. On the basis of the license plate captions, the owner of the car is sent a monthly invoice for the total charge incurred.

Systems with an intrusive nature as the aforementioned have arisen important challenges to the field and have revealed the need of alternative LEZ access control systems which deal with this problematic. This need has led to the appearance of new proposals designed around privacy in which the users' anonymity is treated in a more friendly way.

2 Related Work

In the last years, the improvement of localization and communication technologies have nurtured the evolution of Vehicle Location-Based Services (*VLBS*), which led to the emergence of more flexible and precise *Electronic Road Pricing (ERP)* approaches.

In this way, systems using these technologies are able to more accurately calculate fees on the basis of various factors like the traveled distance or the elapsed time inside the restricted area. In line with that principle, several ERP systems have appeared in the last years [3,4,5,6,7,8]. The main concept behind all these approaches lies in calculating the vehicles' route through the use of their On-Board Unit (*OBU*) which should be equipped with a GPS module. For this purpose, the GPS is periodically tracking the vehicle position and storing its geographical position so as to determinate the corresponding fare according to its activity inside the LEZ. Although these works gather the route information using the same principle, they diverge in the way they manage the gathered data. In [3] and [4] the vehicle's OBU anonymously uploads tuples containing its position to an external server, property of a Service Provider (SP) or similar. Then, during the billing period, the SP calculates the amount to pay for each user on the basis of the uploaded data. Conversely, in the

works presented in [5,6,7,8] is the vehicle's OBU which locally determines the fee to pay and, in each billing period, sends it to the SP as a unique aggregated amount. In order to support this computation and proof its honesty, the OBU also provides cryptographic evidences without disclosing information about its travelled route.

Due to the position tracking client-based nature of the aforementioned ERP schemes, extra anti-fraud measures are required in order to avoid users from intentionally alter the OBU's data gathering, like, for example turning the OBU off or modifying its flow of data. Bearing this in mind, these approaches make use of camera-based checkpoints randomly placed inside the restricted area to counteract these kinds of attacks. This deployment grants the SP the ability of collecting vehicle's location proofs through the recording of license plates. With the objective of proving its honesty, drivers should demonstrate, through different privacy-providing cryptographic mechanisms, that the OBU's collected data is consistent with the SP's license plate recordings. However, the main drawback with this approach is that drivers can spot and intentionally bypass checkpoints in order to continue perpetrating fraud. Increasing the number of checkpoints is usually proposed to detect fraudulent users with a greater probability and overcome this problem. Nevertheless, this measure not only negatively comprises the user's privacy, but also could permit the SP to track vehicles all by himself without needing the OBU's information to estimate users' fees.

Recently, works in [9,10,11] proposed a different privacy-by-design approach which always preserves the users' anonymity unless they show a dishonest behavior. In this way, photos of vehicle's license plates are only taken in case the driver omits, totally or partially, the authentication process with the system infrastructure. In [9] a privacy-preserving protocol based on elapsed time or distance traveled is introduced, offering a fraud control system which, unlike checkpoint-based approaches, does not comprise the privacy of honest drivers. The authors enhance their former protocol in [10] to a multi-fare LEZ scenario which dynamically changes the fare prices according to the traffic density. Later, the work in [11], following the same privacy approach, presents a more lightweight and efficient protocol, which simplifies the vehicles' access data management when the SP is calculating fees during the payment phase. Furthermore, it also proposes using the driver's smartphone instead of the vehicle's embedded OBU as client devices in pursuit of a greater deployability.

As ERP*s*, the previous works are bound to rely on a centralized entity, usually a SP, to gather vehicles' location data that allows to accurately determining the corresponding fee for circulating inside a LEZ at the end of every billing period. This procedure entails great privacy problems that literature works, like the aforementioned, tried to tackle using various strategies.

Furthermore, as a fraud protection, entities involved in the process are required to store signed evidences of each exchange of information they had

made in order to defend themselves from any accusation brought against them; in which case, the whole chain of events should be replicated through these evidences to uncover the dishonest entity.

3 Model of the system

The current LEZ access control systems in the literature rely on centralized entities to keep count of the circulation charges a vehicle's owner has to pay at the end of a billing period. Later, this access log has to be transferred to the entity charging the fees, with all the problems that this poses for users' privacy. Furthermore, every entity involved in the process has to keep evidences of every action to defend himself against errors or fraud attempts.

According to these centralization issues found in the literature due to current ERPs' charging schemes, we propose a new secure and privacy-preserving ERP solution for controlling the access to LEZs, whose fundamental principle is the use of blockchain and smart contracts [12,13] to omit third parties from participating in payment related processes. In this matter, through smart contracts, interactions between users and LEZ's infrastructures are processed as blockchain transactions, thereby permitting the corresponding fee to be automatically calculated and charged from the user's wallet in terms of digital tokens. Under this procedure, entities responsible for registering vehicle accesses and charging their corresponding fees are replaced by a decentralized network, which grants the verifiability, reliability and transparency of the uploaded events. On this basis, there is no need for entities to locally store signed proofs of every interaction they made, as any node belonging to the distributed network can verify the validity of the transaction flow in the blockchain. Taking this approach, however, has no impact on the privacy of honest users and preserves the revocable anonymity property like other works in the literature do.

In this section, the involved actors in the system are introduced first, and then a general view of the protocol is given.

3.1 Actors

Our approach consider the following actors: i) Competent Administration of the LEZ (LA); (ii) Drivers (D); iii) Access Control (AC).

The LA is the entity in charge of managing the LEZ. Among its tasks it is responsible for setting the access rules, the price and deploying the smart contract which manages the accesses and payments of the system. The drivers (D) are the group of users whom the approach is addressed to. Ds' vehicles should be equipped with an tamper-proof OBU with GPS, Bluetooth and 4G embedded in it. The ACs are LEZ infrastructures that control Ds accesses to the restricted area. Among theirs tasks are to act as locations proof generators for vehicles and as a location proof verifiers for the blockchain.
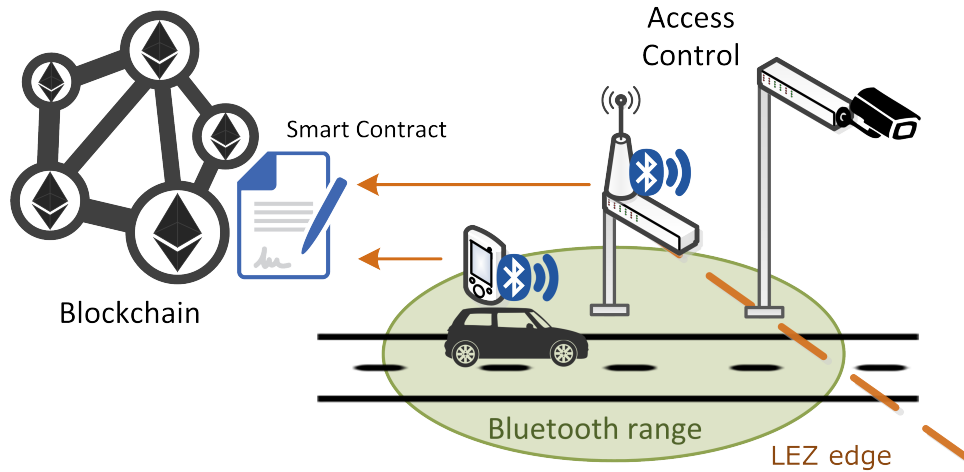
3.2 Our proposal

Fig. 1. Access phase

A general scheme of our LEZ access control system is shown in Figure 1. In this phase of our protocol three entities are involved: the vehicle's OBU, the Access Control infrastructure (AC) and the blockchain network. Before a user can correctly validate her vehicle's entrance into the LEZ, it is required to obtain a digital certificate, which accredits her vehicle's emissions category, from LA entity. Once this certificate is obtained, the vehicle's OBU is able to communicate with the system's infrastructure. On this basis, when the vehicle is about to enter the LEZ, the OBU automatically awakes, through the detection of BLE beacons, and establishes secure connection with the AC. Through the cryptographic protocol the user proves she is driving a valid registered vehicle and agreed an access proof, which will identify the current access in the blockchain. During the whole process, driver's anonymity is maintained through the use of pseudonyms, which can be changed at will to prevent other entities from binding all her accesses. Conversely, if the authentication process does not terminate correctly or somewhat skipped, the AC will take a photo of the vehicle's license plate, revoking the dishonest user's privacy.

Once the authentication with the AC concludes and D obtains her access proof, the payment process is performed as shown in Figure 2. For this, the user's device uploads the agreed access parameters, through smart contract interaction, into the blockchain as a transaction. On the basis these uploaded parameters, the smart contract calculates the amount to pay according to the last uploaded price list and charges the equivalent amount in terms of digital tokens to the user. Later, the involved AC verifies if the transaction has been conducted and the uploaded parameters are the ones agreed during
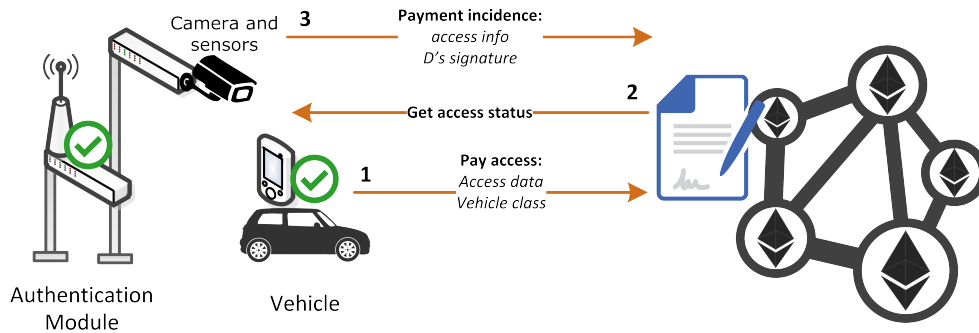
Fig. 2. Payment phase

their authentication process. The AC publishes a claim, which contains its own copy of the access agreement, into the blockchain in case some of the previous premises are not met. The information contained in the claim is enough for the LA to initiate an investigation and disclose the driver's identity.

4 Future Work

As future work, we plan to improve the presented system in three ways: i) achieve a more flexible approach in terms of fee calculation parameters; ii) implement a prototype and the designed smart contract in order to verify its feasibility in realistic scenarios; and iii) enhance the privacy protection for users with a model witch does not require credential renewal to provide untraceability and unlikability.

References

[1] W. H. Organization, UNAIDS, *et al.*, *Air quality guidelines: global update 2005*. World Health Organization, 2006.

[2] G. Santos, "Urban congestion charging: a comparison between london and singapore," *Transport Reviews*, vol. 25, no. 5, pp. 511–534, 2005.

[3] R. A. Popa, H. Balakrishnan, and A. J. Blumberg, "Vpriv: Protecting privacy in location-based vehicular services," 2009.

[4] X. Chen, G. Lenzini, S. Mauw, and J. Pang, "A group signature based electronic toll pricing system," in *2012 Seventh International Conference on Availability, Reliability and Security*, pp. 85–93, IEEE, 2012.

[5] J. Balasch, A. Rial, C. Troncoso, B. Preneel, I. Verbauwhede, and C. Geuens, "Pretp: Privacy-preserving electronic toll pricing.," in *USENIX Security Symposium*, vol. 10, pp. 63–78, 2010.

[6] S. Meiklejohn, K. Mowery, S. Checkoway, and H. Shacham, "The phantom toll-booth: Privacy-preserving electronic toll collection in the presence of driver collusion.," in *USENIX security symposium*, vol. 201, pp. 1–16, 2011.

[7] J. Day, Y. Huang, E. Knapp, and I. Goldberg, "Spectre: spot-checked private ecash tolling at roadside," in *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*, pp. 61–68, ACM, 2011.

[8] F. D. Garcia, E. R. Verheul, and B. Jacobs, "Cell-based privacy-friendly roadpricing," *Computers & Mathematics with Applications*, vol. 65, no. 5, pp. 774–785, 2013.

[9] R. Jardí-Cedó, M. Mut-Puigserver, M. M. Payeras-Capellà, J. Castella-Roca, and A. Viejo, "Time-based low emission zones preserving drivers privacy," *Future Generation Computer Systems*, vol. 80, pp. 558–571, 2018.

[10] R. Jardí-Cedó, J. Castellà-Roca, and A. Viejo, "Privacy-preserving electronic road pricing system for low emission zones with dynamic pricing," *Security and Communication Networks*, vol. 9, no. 16, pp. 3197–3218, 2016.

[11] C. Anglès-Tafalla, J. Castellà-Roca, M. Mut-Puigserver, M. M. Payeras-Capellà, and A. Viejo, "Secure and privacy-preserving lightweight access control system for low emission zones," *Computer Networks*, vol. 145, pp. 13–26, 2018.

[12] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[13] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, pp. 1–32, 2014.

Car localisation for Smart Parking solutions

Daniel Padilla Carrasco *

Department of Computer Engineering and Mathematics,
Universitat Rovira i Virgili, Tarragona, Spain
daniel.padilla@estudiant.urv.cat

Abstract. Smart Parking industry has been demanding new features such as guidance systems, parking surveillance, ticket-less payment, place forecasting, etc. The late improvements in artificial vision and machine learning models, e.g., deep learning, yield robust camera-based solutions for smart parking. One of the top important characteristics of most of these features is car localisation that expresses about the ability to identify each car uniquely. With a vision of proposing a fully automated parking management solutions for smart parking, the first step is car localisation inside parking based on image processing techniques. Camera-based localisation algorithms present an important interest in the field of intelligent video surveillance. In this paper, we study state-of-the-art car localisation, as well as define and analyse the problems of the proposed solutions in outdoor parking scenarios.

1 Introduction

In spite of the increasing popularity that the concept of smart city has gained over the past few years. Smart city concept is mainly about a city that provides ICT-based services in different sectors of activity, in order to mitigate urban challenges, increase efficiency, reduce costs, and enhance the quality of life. Generally, Smart city definition varies in function of city resources, development, and its ability of changing. One of the important things for citizens in smart cities are smart parking systems for monitoring cars in the parking.

Nowadays the simple operation of looking for a free parking spot is becoming a stressful and time wasting task. This problem is especially related to different factors such as:

- Insufficient parking places,
- The large number of accidents due to, the false parking manoeuvres,
- No information about available free spots,
- Vehicle blocked due to parking in unauthorized parking areas,
- Traffic jam due to search of free parking spots,

* Advisors: Hatem A. Rashwan and Domenec Puig

- In the case of massive parking garages, its common for people to forget where they have parked their vehicles

Smart Parking industry use several solutions [9] for parking management like ultrasonic sensor, magnetic sensor or camera based solutions. Using of ultrasonic sensors in indoor parking and magnetic sensors in outdoor parking had been dominating the industry due to simplicity, high precision and sensors low cost. Since camera based solutions introduce more potential features, the industry has recently been using it for several smart parking solutions.

In fact, camera based solutions were introduced with success in indoor scenarios. However, the indoor camera based solution has other challenges that make it less competitive than other solutions. For example, the usual setup of indoor parking restricted the camera height position yield install many cameras to cover up all parking places.

In turn, using the camera-based solutions can be useful in outdoor parking scenarios that face with different external factors like object sizes, illumination and lighting changes. Thus, robust camera based solutions with high performance is necessary to cope with these conditions. Recently with the advance of deep learning models, convolutional neural networks CNN, camera based systems lead to promising results to provide good results from an industry view point perspective [17,9,13,5,10,4].

The more accurate camera-based solutions for parking, the more accurate car localisation techniques. Following that current path, we can find many studies about for detecting the place occupancy and car detection in a parking [17,9,13,5,10,4]. To achieve that, the use of object detection and place occupancy classification techniques will be used for localise, identify and track the cars in the parking.

The CNNs are a machine learning algorithm that uses the local spatial information in an image and learns a hierarchy of increasingly sophisticated features, thus automating the process of feature construction. Recently, CNN-based frameworks have achieved state-of-the-art accuracies in image classification and object detection. A deep CNN (VGGNet-f) has been proposed in [Valipour et al. (2016)] for the application of parking space vacancy identification. The network was fine-tuned to yield a binary classifier with overall accuracy better than 99%. They evaluated the transfer learning ability of the trained classifier on another dataset and reported an accuracy of approximately 95%. A decentralised solution for visual parking space occupancy detection using a deep CNN and smart cameras has been presented in [Amato et al., 2016]. The authors train and fine-tune a miniature version of AlexNet [ref], for binary classification and report an accuracy of 90.7% for the transfer learning process. Similar work has been performed by [Amato et al. (2017)]. The results indicate the achievable accuracy for transfer learning for AlexNet are in the range of 90.52 - 95.60%.

In summary, there is clear evidence in the literature that feature learning by deep CNNs outperform the conventional methods using hand-crafted features for the detection of car parking place occupancy in terms of accuracy, robustness and transfer learning. However, all the CNN-based systems mentioned above fine-tune the existing pre-trained networks, which is an additional training step requiring additional effort.

In this paper, we study and evaluate recent deep learning-based car detection and place occupancy approaches in outdoor parking. The main contributions of the present work are the following:

- Parking place occupancy and car detection detection based on deep learning are assessed and their performance are evaluated.
- A detailed accuracy analysis is performed to identify the parameters that affect the accuracy of the tested frameworks.

2 Car localisation

Car localisation feature can be described as the ability of a smart parking to, given a specific car identification like e.g., its license plate, know exactly where is that car parked. With this feature, one can easily extract other information and features. As examples, it can be known easily, which car places are occupied by looking if there is any car localised over the place (place occupancy), it can guide the user to the parking place with some external application like mobile app (car localisation), or even extract the time between entering a car place and leaving it (precise ticket-less payment).

The car localisation feature then, can be defined as a joint problem where the system has to use several technologies:

- Object detection, to detect every possible semantic object inside the image or video. E.g. Perhaps a light pole may be omitted, but a human could be detected for extra smart parking features(surveillance).
- Object classification, to specifically detect and work with the desired object, cars in this case.
- Object tracking, to follow the location of that car in every frame of the video inside the parking.
- Object identification, being able to tag the car with a unique identifier while inside the parking.

While car localisation as a feature is a complex problem involving several technologies, we can implement this feature step by step. Using the place occupancy feature as a first step to model our neural network as a previous step for car localisation.

2.1 Place occupancy

Place occupancy is a feature to check if a place is occupied or empty inside the parking area. As the location of every place is well-known on the dataset (places are clearly delimited), the common approach for this problem[20,13,5,10,4,2] is to crop the place from the image, resize it to have a shared size and feed it to a binary classifier. The classifier would then output if the place is empty or occupied.

This method does not require the system to detect the objects, since the place ("manually" cropped) is itself the detection. In fact, the output is more direct and concise: occupied or not. However, this method does not contain information about the car, as far as it is within the place or not.

2.2 Car detection

On the other hand, we could take a workaround to get the same results. If we can detect the car using object detection methods, we could know every car location inside the parking. All of them: the ones that are stationed and the ones that are still driving into or driving out. With these information it can be easily mapped, which of the places have a car located within its area.

Obviously, this method is more complex as its inputs are the whole image (normally cropped or resized to the input size of the CNN) and outputs a list of areas (bounding box), one for each car detected. The object detection method is also a more challenging problem since it should be able to detect all objects inside an image, meanwhile the binary classification does only need to check for each manually input area. However, it allow us to compare with the simpler binary classifier method and advance towards the car localisation goal at a slow pace.

3 Experimental Results

As a previous step, we selected different frameworks to be checked and analysed from the current deep learning frameworks for place occupancy and car detection.

For implementing, training and testing these frameworks, both Pytorch[16] and Tensorflow[1] are utilised. However, we will use the Pytorch library in our evaluations. Since, Pytorch has been implemented in a way that it can be easily debugged, hence minimising effort on bugs solving.

In addition, we used the public database, PKLot [3], to assess the performance for place occupancy and car detection. The dataset consists on 12,417 images (around 700,000 parking spaces when cropped) with different illumination conditions depending on the time of the day, which is included in the image file name, and different meteorological statuses (Sunny, Cloudy and
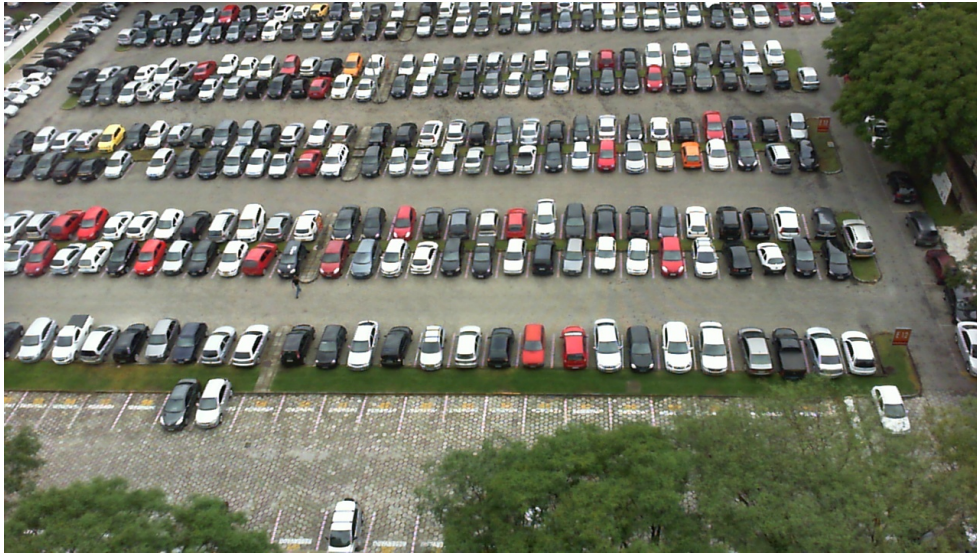
Fig. 1. Example of image on PKLot Dataset

Rainy), which are conveniently split in sub-folders. The images are also split on three subset of images collection from different the different cameras they were taken. One of the cameras giving near-aerial view shots, and high angle views for the other two.

3.1 Place occupancy

On this experiment, we crop every car and non-car places from our dataset to train the models for binary classification (i.e., empty or occupied).

For the experiment, we selected ResNeXt[21], a network for object classification. The ResNeXt is also trained with cardinality = 4, widen factor = 4, base width = 32 and depth = 29. Besides, we used the stochastic gradient descent (SGD) as an optimisation method with cross entropy as a loss function. We achieved an accuracy of about $99,00\%$ on validation set at epoch 150. Figure 2 shows the validation error rate percentage (as Val Error in table) and the loss error (as Loss Error in table) of the trained ResNeXt.

In addition, We tried the Nesterov momentum as an optimiser to check if the convergence will be faster. And although, we did not get quicker results, but the convergence was improved a little bit as shown in Table 1.

Optimisation	Val Error	Loss Error
SGD	0,09969	5.5197*10-3
Nesterov momentum	0.08393	4.2256*10-3

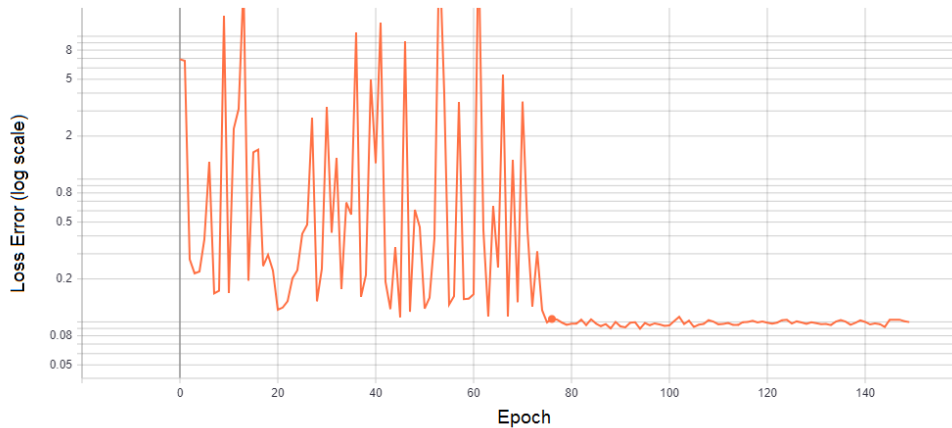**Table 1.** Optimisation method at epoch 150

Fig. 2. Loss error of the trained ResNeXt network with 150 epochs.

We also checked another model called PyramidNet[8] in two architecture: low profile and medium profile to reduce the computational cost. Where:

- Low Profile: Trained PyramidNet with parameters Base width = 8 and alpha=30
- Mid Profile: Trained PyramidNet with parameters Base width = 16 and alpha=300

However, as shown in Table 2, the ResNeXt with adding Batch normalisation function in the end of each layer yields better results than the PyramidNet architectures. Figure 3 shows different results of loss error with different architectures of the PyramidNet model showing that the ResNeXt with Batch Normalisation and Dropout in last layer of the network provides the lowest loss error among the tested models.

Net	Val Error	Loss Error
ResNeXt	0,09969	5.5197*10-3
PyramidNet Low profile	0,1979	8.0803*10-3
PyramidNet Mid profile	0.1441	6.4133*10-3

Table 2. Comparison between ResNext trained model and PyramidNet trained models

3.2 Object Detection

We should also checked the performance on the SSD network [15] for object detectotion. For training, we use SSD with VGG-16[19] pre-trained model in order to take advantage of the low level features extracted by the first layers
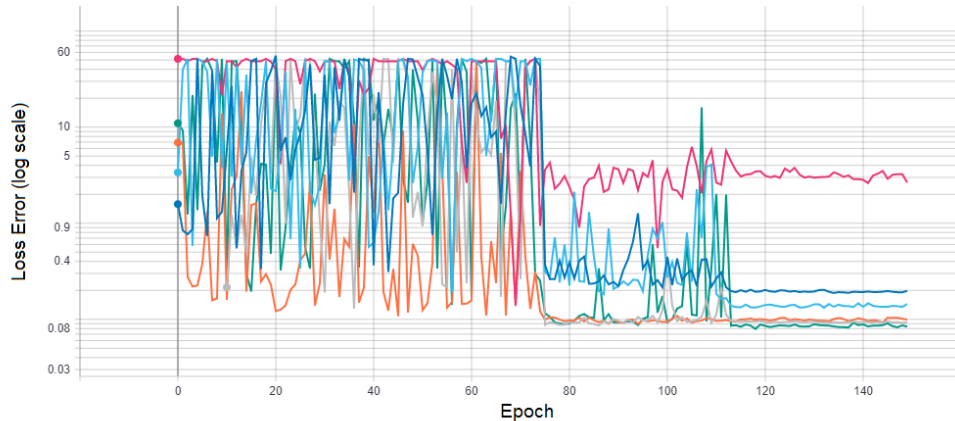
54

Fig. 3. Different training evolution (log scale). (Purple) Batch Normalisation plus Dropout, (Dark Blue) PyramidNet Low profile, (Light Blue) PyramidNet Mid profile, (Orange) ResNeXt, (Gray) Batch Normalisation and Dropout in last layer of ResNeXt as in [14], (Turquoise) Batch Normalisation with all layers of ResNeXt.

of the network. Note that VGG-16 is pre-trained with a ten classes dataset, including car. For a first experiment, we trained the SSD model with VOC-2007[7] dataset and test the results with PKLot Dataset. On a second, we replaced some car images from VOC07 for some of the PKLot dataset images and fine-tuned with the VGG-16 pretrained model. As shown in Table 3, the results are not accurate, since car perspective is so much different from the samples used for training the VGG-16 layers. Finally, we fine-tuned the pre-trained VGG-16 model with just PKLot dataset. For this last experiment we adapted the annotations of the PKLot database to match those expected from the SSD model using the following steps:

- Cropping (300x300) around a randomly selected place.
- Include, in the image annotations, the remaining places inside the image as in 4.
- Modify annotations file so that, even though there are still 21 classes for training, only 1 class is feed to the model.

However, as shown in Table 3, we can not reach similar values to SSD paper for VOC07.

4 Discussion and future work

Regarding the place occupancy experiment, we can reach an accuracy of +99% on the validation set of the PKLot dataset. That is a very high performance that can be beneficial for any industry solution. However, as stated before, it will get extremely difficult to extend that solution to a car localisation

Fig. 4. Examples of the Results of the SSD model with the PKLot datset.

Dataset training	mAP
VOC-07	0,7749
Edited VOC-07	0,0065
VOC-PKLot	0.3334

Table 3. (VOC-07) Trained base SSD with VOC-2007 dataset. (VOC-07 edited) Replaced VOC-2007 car images with PKLot resized images. (VOC-PKLot) VOC Annotations for PKLot and 1 class input.

problem, where we need to identify each car within a large number of them. Thus, we need to depend object detection in order to be able to track each car. However, as shown in Table 3, the results are still far from being competitive with a simple place occupancy experiment. Since, the PKLot dataset contains of very tiny samples of cars. The SSd model can not able to provide good results. Therefore, we need to work on developing a more accurate object detection algorithm.

In the case of PKLot, the images contains tiny samples of the cars (i.e., about 30x30 pixels) inside the whole image that adds a difficulty to the detection algorithm. For this reason we are currently working on adapting the SSD and other detection networks to be useful for tiny object detection in a scenario of crowd cars [6,12,11,18]. For example, a multi-scale SSD model, called MSSD, was proposed in [6] that can be suitable for better detection with these changes in scale.

5 Disclosure

This research is sponsored by Quercus Technologies S.L. and may lead to the development of products which may be licensed to it, in which I have a financial interest.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, and et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] Debaditya Acharya, Weilin Yan, and Kourosh Khoshelham. Real-time image-based parking occupancy detection using deep learning. In *Research@Locate*, 2018.

[3] P. Almeida, L. S. Oliveira, E. Silva Jr, A. Britto Jr, and A. Koerich. Pklot a robust dataset for parking lot classification. volume 42, page 11, 2015.

[4] Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Carlo Meghini, and Claudio Vairo. Deep learning for decentralized parking lot occupancy detection. *Expert Systems with Applications*, 72:327–334, 2017.

[5] Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, and Claudio Vairo. Car parking occupancy detection using smart camera networks and deep learning. In *Computers and Communication (ISCC), 2016 IEEE Symposium on*, pages 1212–1217. IEEE, 2016.

[6] Lisha Cui. MDSSD: multi-scale deconvolutional single shot detector for small objects. *CoRR*, abs/1805.07009, 2018.

[7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

[8] Dongyoon Han, Jiwhan Kim, and Junmo Kim. Deep pyramidal residual networks. *IEEE CVPR*, 2017.

[9] K. Hassoune, W. Dachry, F. Moutaouakkil, and H. Medromi. Smart parking systems: A survey. In *2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA)*, pages 1–6, Oct 2016.

[10] Meng-Ru Hsieh, Yen-Liang Lin, and Winston H. Hsu. Drone-based Object Counting by Spatially Regularized Regional Proposal Network. *arXiv e-prints*, page arXiv:1707.05972, Jul 2017.

[11] Guo X. Hu, Zhong Yang, Lei Hu, Li Huang, and Jia M. Han. Small Object Detection with Multiscale Features. *International Journal of Digital Multimedia Broadcasting*, 2018:10, 2018.

[12] Peiyun Hu and Deva Ramanan. Finding tiny faces. *CoRR*, abs/1612.04402, 2016.

[13] K. Kaarthik, A. Sridevi, and C. Vivek. Image processing based intelligent parking system. In *2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE)*, pages 1–4, April 2017.

[14] Xiang Li, Shuo Chen, Xiaolin Hu, and Jian Yang. Understanding the Disharmony between Dropout and Batch Normalization by Variance Shift. *arXiv e-prints*, page arXiv:1801.05134, Jan 2018.

[15] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In *ECCV*, 2016.

[16] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[17] P. Ramaswamy. Iot smart parking system for reducing green house gas emission. In *2016 International Conference on Recent Trends in Information Technology (ICRTIT)*, pages 1–6, April 2016.

[18] Yun Ren, Changren Zhu, and Shunping Xiao. Small object detection in optical remote sensing images via modified faster r-cnn. *Applied Sciences*, 8:813, 05 2018.

[19] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[20] S. Valipour, M. Siam, E. Stroulia, and M. Jagersand. Parking-stall vacancy indicator system, based on deep convolutional neural networks. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pages 655–660, Dec 2016.

[21] Saining Xie, Ross Girshick, Piotr Dollr, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *arXiv preprint arXiv:1611.05431*, 2016.

NES: Nash Equilibrium Scheme for broadcasting protocols in VANETs

Assia NAJA *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
naja.assia@gmail.com

1 Introduction

Vehicular Ad-hoc Networks (VANETs) are a category of Mobile Ad-hoc Networks (MANETs) characterized by frequent topology changes. VANETs have many promising applications such as improving vehicle drivers' safety, and decreasing accidents. Vehicles communicate by broadcast, which needs to be efficient, especially in congested areas. An effective rebroadcasting protocol has to be designed carefully in order to maximize reachability while minimizing delay and the number of rebroadcasts. However, prior work typically minimizes either congestion or latency at the expense of reachability and the number of rebroadcasts. In this work, we present *Nash Equilibrium Scheme (NES),* a rebroadcast algorithm that outperforms prior work on reachability, number of rebroadcasts and delay, especially in congested areas. *NES* attains these goals by considering key environmental factors such as number of received messages and distance to event. We evaluate *NES*'s performances through NS2 simulations.

2 NES: Nash equilibrium scheme: Overview and analysis

In this section, we present a novel protocol which allows the right decision to be made about whether or not to forward the message.

The algorithm is presented in Algorithm 1. First, each vehicle forwards received packets if they have not been received before by checking the packet ID. Otherwise, already received packets are automatically discarded. Second, we assume that all nodes have an identical transmission range.

2.1 Nash equilibruim analysis

The game G = $\{N, A_{i,i \in N}, U_{i,i \in N}\}$ is formulated as a modified volunteer's dilemma, where $A_1, A_2, ..., A_n$ are the sets of pure strategies of players (Vol-

---

* PhD advisor: Pr. Domenec Puig

Algorithm 1 *Pseudo code of the proposed Nash Equilibrium Scheme NES*

 1: **Input**: Received a broadcast message M
 2: **function** DECISION(message M)
 3:     **Step 1:** Upon receiving a message M for the first time
 4:     **if** (M is received before) && (M is forwarded before) **then**
 5:         Discard M and exit;
 6:     **else**
 7:         **Step 2:** Initialize a counter $nm=1$;
 8:         Extract (X_a, X_b) /Coordinates of the receiver;
 9:         Extract (X_0, X_0) /Coordinates of the incident;
10:         Calculate d_i;
11:         **if** $d_i >$ D **then**
12:             Discard M and exit;
13:                                     ▷ (D is a specified threshold)
14:         **end if**
15:         **Step 3:** Set a random time T $\in [0, T_{max}]$;
16:         **Step 4**: Wait until T expires
17:         **if** the same packet is received again **then**
18:             Increase the counter nm by 1;
19:         **end if**
20:         **Step 5:** When T expires;
21:         Compute C_i and p_i;
22:         Generate a random number r=rand(0,1);
23:         **if** r $< p_i$ **then**
24:             Forward M and exit;
25:         **else**
26:             Discard the packet and exit;
27:         **end if**
28:     **end if**
29:     **Return** Retransmitting or discarding M
30: **end function**

---

unteers); and $U_1, U_2, ..., U_n$ are the player (volunteer) payoffs. For each VN$_i$, actions are $A_i = \{F, NF\}$, which means that each vehicle chooses whether or not to rebroadcast the received message. Then, we separate the strategy of each VN$_i$, a_i, and its opponents by using the following notation (a_i, a_{-i}). VN$_s$ participate in the cooperative forwarding (F) according to a probability of volunteering p_i, while others may prefer not to participate (NF) with 1-p_i. This probability can be calculated using parameters evolved in the game. Each VN$_i$ has a forwarding cost C$_i$ associated with message broadcasting. We consider that the benefit is unitary (B=1). And we assume that $0 < C_i < 1$ (which means that $B > C_i$), so each node will prefer to participate in the game if no other player does. The outcome of each VN is as follows:

$$U_i(a_i, a_{-i}) = \begin{cases} -r_i & P = \emptyset \\ 1 & i \notin P \ \& \ P \neq \emptyset \\ 1 - C_i & Otherwise \end{cases} \quad (1)$$

60

r_i is the regret of a VN when no player cooperates. Moreover, each VN_i obtains a regret of $-r_i \leq 0$. Table 1 presents the outcomes of two heterogeneous VN_i (different forwarding costs) in the regular way:

Table 1: Normal form of the game

	F	NF
F	1-C_i,1-C_j	1-C_i,1
NF	1, 1-C_j	$-r_i,-r_j$

2.2 The probability of volunteering

The definition of a fully mixed Nash equilibrium [1] is as follows: "A fully mixed strategy Nash Equilibrium specifies a fully mixed strategy $p_i^* \in\ ]0,1[$, such that for each player i:

$$U_i(p_i^*, p_{-i}^*) \geq U_i(p_i, p_{-i}^*) \quad \forall p_i \in]0,1[\tag{2}$$

We denote by $U_{iF}(p_i)$ the utility of the VN_i when it chooses to participate in the rebroadcasting game, and $U_{iNF}(p_i)$ its utility when its choice is NF. Thus, the expressions of $U_{iF}(p_i)$ and $U_{iNF}(p_i)$ are as follows:

$$U_{iF}(p_i) = 1 - C_i \tag{3}$$

$$U_{iNF}(p_i) = (1 - (1 - p_i)^{\frac{1}{N-1}}) - r_i(1 - p_i)^{\frac{1}{N-1}} \tag{4}$$

After some algebraic calculations, we obtain the equation of the rebroadcasting probability p_i, which is expressed as follows:

$$p_i = 1 - (\frac{C_i}{1 + r_i})^{\frac{1}{N-1}} \leq 1$$

2.3 Expression of the cost

the expression of the cost C can be expressed as follows:

$$C_i = f(d_i) * f(t_i) \tag{5}$$

The distance and node speed are two independent events. Therefore, the cost of participating in the game is achieved by making a product of the functions $f(d_i)$ and $f(t_i)$. The function $f(d_i)$ can be expressed as:

$$f(d_i) = \begin{cases} \frac{d_i}{D} & d_i \leq D \\ 1 & d_i \geq D \end{cases} \tag{6}$$

the time cost function $f(t_i)$ can be calculated as:

$$f(t_i) = \begin{cases} 1 - \frac{t_{min}}{t_i} = 1 - \frac{s_i}{S_{max}} & t_i \geq t_{min} \\ 0.1 & t_i \leq t_{min} \end{cases} \tag{7}$$

Where the current vehicle has the speed s_i and each vehicle has a maximum speed S_{max}.

3 SIMULATION PARAMETERS

The common simulation input parameters are illustrated in Table 2.

Simulation Parameters		
Scenario	Type	Speed
	Grid-map	60 km/h
Propagation model	Two-ray ground	
MAC and phy	Type	TR	Power
	802.11	300m	17 dBm
Packet size	500 bytes	
Simulation time	1000 secondes	
# nodes	50, 100 and 150 nodes	
# trials	20	

Table 2: Summary of the simulation parameters

4 SIMULATION RESULTS

5 CONCLUSION

In this work, we present an adaptive probabilistic algorithm based on game theory in VANETs, and we evaluated existing protocols using NS2 in several scenarios that varied node density and speeds. NS2 simulation results prove that NES can keep a balance between reachability and rebroadcast efficiency by generating fewer rebroadcasts or more saved rebroadcasts, than other algorithms.
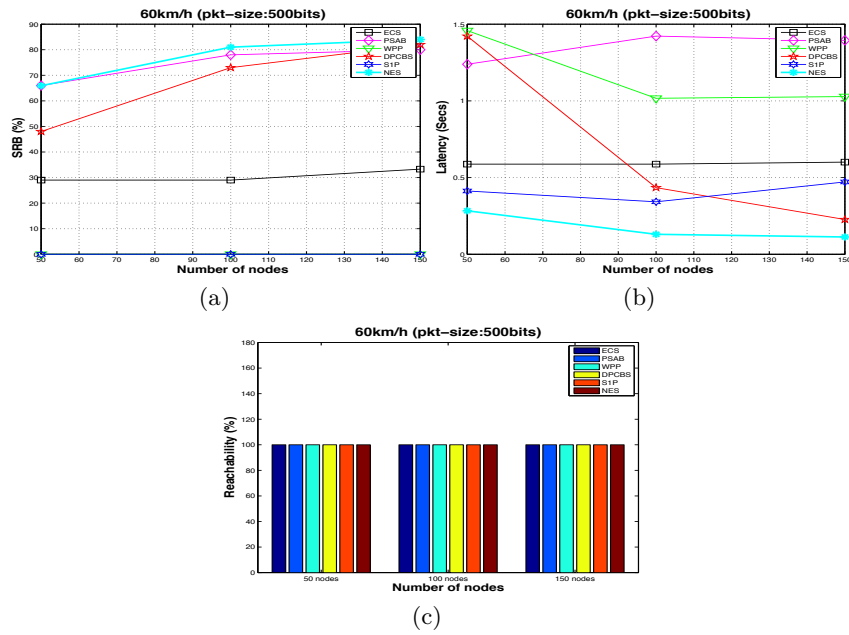
Fig. 1: Latency for 60 km/h in a grid.

References

[1] EL HAMMOUTI, Hajar, SABIR, Essaid, ECHABBI, Loubna, et al. Identifying a volunteer-like dilemma in cooperative sensing-empowered cognitive radio networks. In : 2016 IEEE International Conference on Communications (ICC). IEEE, 2016. p. 1-7.

Process Mining and Spaghetti Business Processes: An Uneasy Relationship

Edgar Batista [*]

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
`edgar.batista@urv.cat`

1 Introduction

As a result of a more demanding population, the achievement of effective, cost-efficient and sustainable models of society services, such as healthcare or transportation, is one of the main challenges nowadays. To meet this goal, the information and communication technologies (ICT) have played a key role thanks to their rapid development, and they have opened the door to novel opportunities. For instance, within the healthcare sector, this fact has promoted the *smart health* paradigm, aiming at providing health services founded on the use of network capabilities and the sensing infrastructure of context-aware environments [13]. Advancements on channel wireless characterisation in medical scenarios [5], promoting healthier lifestyles with recommender systems [6], and automatically detecting wandering behaviour experienced by people with dementia [3] are only a taste of the opportunities resulting from converging ICT, context-awareness and healthcare, which lead to valuable knowledge [12].

Within an organizational context, services provided to the society can be understood as *business processes*. A business process comprises a group of related activities aiming at fulfilling a certain organizational goal. Even though managing business processes has great benefits for organizations (*e.g.* visualize real process executions, identify bottlenecks within the processes. . .), this can be extremely hard, especially in organizations with a large number of business processes that are being executed concurrently. To facilitate the management task, the execution of business processes leaves traces in the form of *events*, and they are stored in the so-called *event logs*. Each event consists of multiple attributes, such as a unique identifier, a process execution case identifier, an activity, and a timestamp. With the aim to discover, monitor and improve real processes by extracting knowledge from event logs readily available in today's systems, the *process mining* research discipline emerged in the beginnings of the 2000s [1].

---

[*] PhD advisor: Agusti Solanas

Despite the excellent process mining tools, sometimes the success of extracting valuable knowledge does not depend on the process mining techniques and algorithms, but on the nature of the business processes themselves. In contexts characterized by their high degree of complexity and dynamism, such as healthcare, where process mining is gaining importance [4], the fact of discovering business processes with no (or little) structure is commonplace. This kind of business processes are popularly known as *spaghetti processes*. Spaghetti-like processes pose serious challenges for analysis because their lack of structure makes them hard to understand and comprehend (cf. Figure 1). It is worth noting that these processes are not incorrect and, in fact, they are feasible when reflecting the reality from event logs very accurately (*e.g.* representing all process variants, process executions are driven by experience or intuition. . .). For instance, medical treatments (understood as business processes) might behave differently between patients for a given disease. Also, the quality of the event logs is another important factor that could lead to spaghetti processes. Since real environments are far from ideal, event logs can be noisy (*i.e.* containing exceptional or infrequent behaviour) or incomplete (*i.e.* not representing all the behaviour) [2].

Keeping process models in a simple fashion when describing a well-representative behaviour from the event log is not straightforward. In this article, we explore a novel technique that, by design, simplifies process models during the execution of a process mining discovery algorithm, named `skip miner`. The rest of the article is organized as follows. Section 2 describes current techniques for simplifying process models from the state of the art. Then,
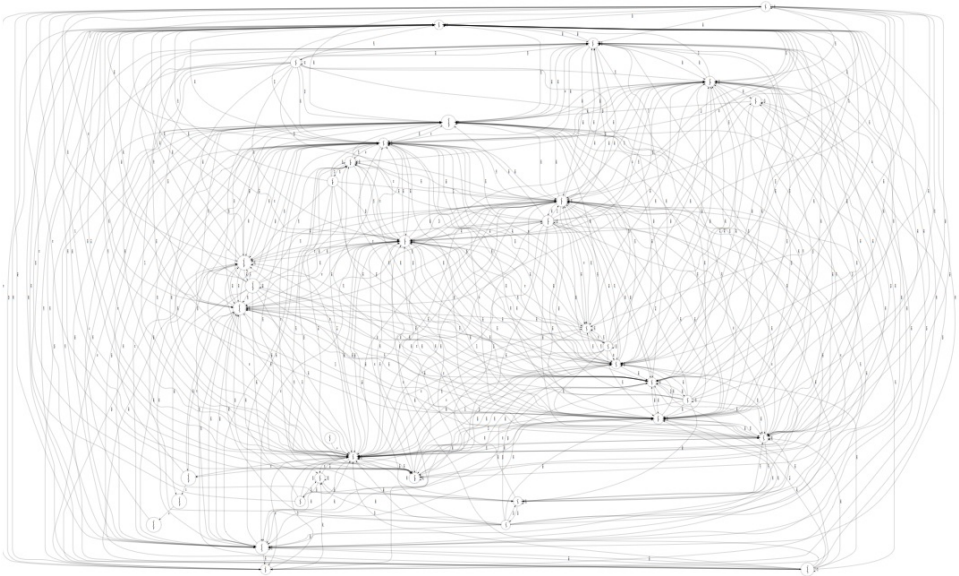

Fig. 1. Example of a spaghetti-like business process model.

our `skip miner` approach is introduced in Section 3, and later evaluated and compared in Section 4. Finally, Section 5 closes the article.

2 Related work

One the most common strategies to bring some structure to spaghetti-like process models and make them easier to comprehend consists in approaching some simplification technique at the time of discovering the process models.

On the one hand, the simplification procedure can even begin before the discovery of the process models themselves, by simplifying directly the event information from the event logs. In general, this idea attempts to represent models with the most frequent behaviour only, and abstracting it from detailed or infrequent behaviour that, in large, results into spaghetti process models. For instance, considering only events whose activity is frequently repeated along the event log is a common (and fast) way to filter out events that rarely happen. Also, the trace clustering method [14] aims at splitting the event log into homogeneous subsets of execution cases and, for each of them, represent it adequately by an independent process model, which will be simpler compared to the process model depicting the entire behaviour of the event log.

On the other side, there exist methods for simplifying process models after their discovery from the event logs. For instance, the removal of activities and/or edges with little importance within the overall process model, by means of thresholds is a common way to simplify process models from irrelevant behaviour. Fuzzy mining [8], one of the most popular process mining discovery algorithms, allows the aggregation of highly-correlated nodes into clusters and the filtering of irrelevant edges. Last but not least, the inductive miner algorithm [9] also applies multiple filters throughout the discovery of process models, by removing infrequent cases as well as infrequent edges.

3 The skip miner method

With the aim to achieve structure and understanding within process models, this article introduces `skip miner`, a novel method for simplifying process models during the discovery procedure itself. This method is different from classical simplification schemes in two main aspects: (i) no definition of simplification-related thresholds is required because the simplification strategy is part of the discovery procedure, and (ii) not all the events are considered, thus skipping/omitting certain events that are likely to transform the model into a spaghetti-like model. The details of the `skip miner` method are described below.

An event log L encompasses a set of traces $t_1, t_2, \ldots, t_p$ representing different execution cases of a certain business process. Each trace t has multiple events $e_1, e_2, \ldots, e_n$ that indicate, among others, the activities realized

chronologically within that process. In this article, the processes represent the activities of the events, so in this context, talking about events or talking about activities is analogous. Although there are several ways to represent a process model (*e.g.* BPMN, petri net...), this method represents it with a weighted directed graph $G = (V, E)$, where V represents the activities of the events as vertices/nodes, and E represents the transitions between two vertices from V with a certain probability of occurrence.

To obtain a complete and detailed view of the business process and how it was executed, the graph G should represent all the transitions between consecutive pairs of events belonging to the same process execution case. Nevertheless, scenarios where the event log is very large or contains traces with very diverse events, the discovered graphs are likely to be spaghetti. To avoid this issue, the `skip miner` method does not consider all the transitions between consecutive pairs of events, but only some of them, whilst the rest of the events (as well as the transitions from/to these events) are skipped from the representation. The number of events to skip when considering a transition between two events comes determined by a numeric parameter *skip* δ ($\delta \geq 0$). As an example, given a trace $t = \langle e_1, e_2, e_3, e_4 \rangle$, if the `skip miner` method decides to skip one event ($\delta = 1$) in e_1, then event e_2 is skipped and the transitions are $e_1 \rightarrow e_3$ and $e_3 \rightarrow e_4$. Indeed, this method can generalize to never skipping events, thus considering all the events from the event log when $\delta = 0$. Figure 2 illustrates the general idea of the `skip miner` method when skipping events from an event log according to the δ parameter, when constructing the transitions between events that will be represented in the graph.

The fact that the `skip miner` method skips certain events with the goal to simplify the resulting process models arises a second question: when should events be skipped? Undeniably, skipping events distorts the original behaviour of the process model, so there must be a trade-off between simplification and information loss. The first idea that comes to mind is deciding skipping events at each step. For instance, given a trace $t = \langle e_1, e_2, e_3, e_4, e_5, e_6, e_7 \rangle$, if $\delta = 1$, then the resulting transitions are $e_1 \rightarrow e_3$, $e_3 \rightarrow e_5$ and $e_5 \rightarrow e_7$; if $\delta = 2$, then the resulting transitions are $e_1 \rightarrow e_4$ and $e_4 \rightarrow e_7$, and so on. *A priori*, this design decision simplifies notably the number of nodes and edges of the resulting graph, but a significant information seems to be lost at first sight, which increases together with the value of δ. Also, this decision has another important drawback: the fact of skipping events does not consider their type neither their information. In a graph (or a process model, in general), not all nodes and edges have the same importance, so the fact of skipping some events can have different impact. With the aim to balance the trade-off between simplifying the process model and minimizing the information loss, a heuristic, described below, has been designed and integrated in the logic of the `skip miner` method.
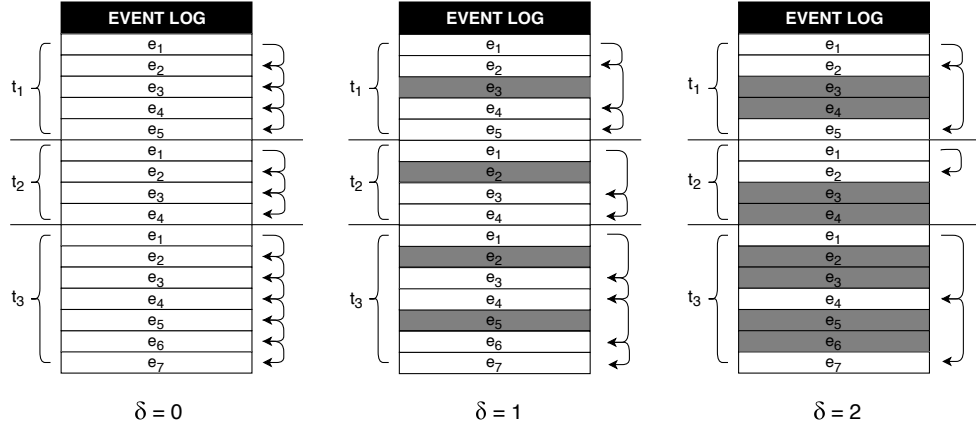
Fig. 2. General behaviour of the skip miner method and the functionality of the parameter δ.

As stated before, in a graph G, nodes/vertices represent activities that have been executed in the events, and edges represent transitions between activities. In order to skip events (which is analogous as skipping activities), each node has a *skip probability* ε, representing the probability to skip the activities of the next δ events. Formally, the skip probability of a node n, $\varepsilon(n)$, is bounded between 0 and 1, where 0 means that there is no probability of skipping events (so, the next event is always considered), and 1 means that the following events are skipped. Given a node n, its value ε comes determined by Equation 1.

$$\varepsilon(n) = 1 - \frac{1}{\#succ(n)} \tag{1}$$

According to this, the more successors (*i.e.* different activities of its subsequent events) a node has, the more probability to skip events. In spaghetti-like process models, it is highly common to find several paths after visiting a specific node/activity, which is the result of different behaviours experienced during the different process executions. The main idea behind this heuristic is, in fact, trying to minimize the complexity of these kind of nodes, called *hub nodes*, which are those nodes with a high skip probability. This heuristic assumes that not all the subsequent nodes/activities from a hub node are equally important, so removing some of them is a way to simplify the graph model. Also, in case of not skipping the subsequent events after a hub node, it is more probable that the following events to represent in the graph are those with more frequency of appearance within the event log, so maintaining the most frequent edges after a hub node. In contrast to hub nodes, there are *direct nodes*, which are nodes whose subsequent activity is always the same. As it can be observed in the skip probability, these nodes always have $\varepsilon = 0$,

so preventing skipping the subsequent activity that, at the end, does not provide any benefit from the simplification point of view and, indeed, it would introduce additional noise that increases the information loss.

4 Experiments

This section presents some experimental results of the proposed `skip miner` method using a real event log from a healthcare institution in the region of Tarragona. More specifically, this event log contains thousands of events detailing the activities conducted by the practitioners during the ongoing of the medical treatments for their patients. In the event log, each event is defined by six attributes: an event identifier, a trace identifier, a doctor identifier, a patient identifier, the name of the activity executed, and a timestamp.

To evaluate the `skip miner` method in this article, we only focus on the discovery of the process model, in form of graph, of a particular doctor, whose name is not disclosed for privacy reasons. To do this, we extract the events from the entire event log associated to this doctor. As a result, we obtain a total of 827 events, associated to 279 different traces (each of them is an independent treatment for a patient), in which the doctor has realized up to 27 different activities. Before testing our method, we obtain the original behavioural graph model G of the doctor, by considering all the activities and all the transitions between consecutive events (*i.e.* analogous as executing the method `skip miner` with $\delta = 0$). Then, we execute the `skip miner` method with different values of the parameter $\delta = \{1, 2, 3, 4, 5\}$ to observe how the resulting graphs differ regarding the original graph. Also, we have implemented another method from the state of the art that simplifies spaghetti process models: considering only events whose activity is frequently repeated along the event log, hereafter named the `activity filtering` method for simplicity reasons. More specifically, the modeling of the graphs in the `activity filtering` method depends on a parameter μ, which defines the proportion of traces that an activity must appear in order to be considered. This method is executed with different values of the parameter $\mu = \{0.01, 0.03, 0.05, 0.08, 0.1, 0.15\}$, where $\mu = 0.01$ means that an activity must appear in, at least, the 1% of the traces, and so on.

The results in this study are used to analyse two main facts. On the one hand, the simplification of the process graphs by means of the number of nodes, number of edges, and the proportion of edges per node in the graph. On the other hand, the distortion of the process graphs in regards to the original graph, by means of similarity measures, such as the Vertex Edge

Overlap (VEO) [10], Vertex Ranking (VR) [10], network-traffic Graph Edit Distance (nt-GED) [7] and Weight Distance (WD) [11][2].

Table 1. Results of both the `skip miner` and the `activity filtering` methods.

		Graph properties			Similarity			
		#nodes	#edges	edges/node	VEO	VR	nt-GED	WD
Original graph G		27	117	4.33	-	-	-	-
	1	25	107	4.28	0.72	0.86	20.4	0.69
	2	25	86	3.44	0.69	0.84	20.6	0.73
Skip miner (δ)	3	21	66	3.14	0.58	0.68	28.7	0.81
	4	19	46	2.42	0.50	0.62	32.2	0.86
	5	18	43	3.39	0.49	0.56	34.5	0.87
	0.01	19	105	5.53	0.91	0.64	16.1	0.15
	0.03	15	97	6.47	0.84	0.51	25.1	0.27
Activity filtering (μ)	0.05	11	80	7.27	0.72	0.39	34.9	0.47
	0.08	8	52	6.5	0.58	0.28	39.4	0.65
	0.1	7	40	5.71	0.49	0.25	41.3	0.73
	0.15	5	25	5	0.33	0.18	45.7	0.87

The results of these experiments are shown in Table 1. Regarding the `skip miner` method, the increase of the parameter δ simplifies the structure of the graph model, since fewer nodes and edges are considered. In parallel, this affects the quality of the process model, which steadily differs from the original graph, according to the different similarity measures. A similar effect is shown when executing the `activity filtering` method and increasing the value of parameter μ. Yet, it is worth mentioning that the `activity filtering` method has a higher impact when simplifying the process models, since the pace of discarding nodes is much faster than the `skip miner` method. In spite of this, this causes that the resulting graphs are, in general, more distant from the original graph, rather than the graphs resulting from the `skip miner`. In this experiment, we can observe that the `activity filtering` method rapidly reduces the number of nodes and keeps only the most frequent activities, but this has a significant impact on the topology of the graph as well as on the information loss. On the other hand, the `skip miner` method attempts to balance the trade-off between this two facts, and only reduces the complexity of the graph when needed, as well as considering the minimization of the information loss.

---

[2] VEO and VR are bounded between $[0-1]$, where 1 indicates total similarity; nt-GED is bounded between $[0-\infty]$, where 0 indicates total similarity; and WD is bounded between $[0-1]$, where 0 indicates total similarity.

5 Conclusions

The need for efficient and sustainable service models has raised awareness on the correct management and monitoring of the execution of the business processes within organizations. To accomplish this arduous task, research on process mining has made great advances in recent years thanks to the increasing attention that has received from the research community. Among others, the discovery of the real executions of business processes using event data from organizational event logs is one of the main uses, but it sometimes leads to extremely unstructured and complex process models that are difficult to comprehend and, in consequence, to acquire valuable knowledge from them. This kind of models are popularly known as spaghetti process models.

Providing structure to spaghetti-like process models is not straightforward, so it is important to develop novel strategies that contribute to the better understanding of process models. This article has presented a novel algorithm, called `skip miner`, for discovering process models represented as graphs, aiming at simplifying the complexity of the resulting models by skipping certain events from the event logs. This method has been tested with a real event log from a large healthcare institution in the region of Tarragona. Besides, the experimental results have also been compared with another simplification method from the literature.

Further work will concentrate on conducting more exhaustive analysis of the proposed `skip miner` method with a larger set of process models and event logs in order to validate the feasibility and robustness of the method.

References

[1] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes.* 1st edition, Springer, Berlin Heidelberg, 2011.

[2] W.M.P. van der Aalst, A. Adriansyah, A.K. Alves de Medeiros, F. Arcieri, T. Baier, T. Blickle, J.C. Bose, P. van den Brand, R. Brandtjen, J. Buijs, A. Burattin, J. Carmona and others. Process Mining Manifesto. *Proc. 9th. Int. Conference on Business Process Management*, 169–194, Clermont-Ferrand, France, 2011.

[3] E. Batista, F. Borras, F. Casino and A. Solanas. A study on the detection of wandering patterns in human trajectories. *Proc. 6th. Int. Conference on Information, Intelligence, Systems and Applications*, 1–6, Corfú, Greece, 2015.

[4] E. Batista and A. Solanas. Process Mining in Healthcare: A Systematic Review. *Proc. 9th. Int. Conference on Information, Intelligence, Systems and Applications*, 1–6, Zakynthos, Greece, 2018.

[5] F. Casino, L. Azpilicueta, P. Lopez-Iturri, E. Aguirre, F. Falcone and A. Solanas. Optimized Wireless Channel Characterization in Large Complex Environments by Hybrid Ray Launching-Collaborative Filtering Approach. *IEEE Antennas and Wireless Propagation Letters*, 16:780–783, 2017.

[6] F. Casino, C. Patsakis, E. Batista, F. Borràs and A. Martínez-Ballesté. Healthy Routes in the Smart City: A Context-Aware Mobile Recommender. *IEEE Software*, 34(6):42–47, 2017.

[7] P. Dickinson and M. Kraetzl. Novel approaches in modelling dynamics of networked surveillance environment. *Proc. 6th. Int. Conference on Information Fusion*, 302–309, Cairns, Australia, 2003.

[8] C.W. Günther and W.M.P. van der Aalst. Fuzzy Mining - Adaptive Process Simplification Based on Multi-perspective Metrics. *Proc. 5th. Int. Conference on Business Process Management*, 328–343, Brisbane, Australia, 2007.

[9] S.J.J. Leemans, D. Fahland and W.M.P. van der Aalst. Discovering Block-Structured Process Models From Event Logs Containing Infrequent Behaviour. *Proc. 11th. Int. Conference on Business Process Management*, 66–78, Beijing, China, 2013.

[10] P. Papadimitriou, A. Dasdan and H. Garcia-Molina. Web graph similarity for anomaly detection. *Journal of Internet Services and Applications*, 1(1):19–30, 2010.

[11] P. Shoubridge, M. Kraetzl, W. Wallis and H. Bunke. Detection of abnormal change in a time series of graphs. *Journal of Interconnection Networks*, 3(1–2):85–101, 2002.

[12] A. Solanas, F. Casino, E. Batista and R. Rallo. Trends and challenges in smart healthcare research: A journey from data to wisdom. *Proc. IEEE 3rd. Int. Forum on Research and Technologies for Society and Industry*, 1–6, Modena, Italy, 2017.

[13] A. Solanas, C. Patsakis, M. Conti, I.S. Vlachos, V. Ramos, F. Falcone, O. Postolache, P.A. Pérez-Martínez, R. Di Pietro, D.N. Perrea and A. Martínez-Ballesté. Smart health: A context-aware health paradigm within smart cities. *IEEE Communications Magazine*, 52(8):74–81, 2014.

[14] M. Song, C.W. Günther and W.M.P. van der Aalst. Trace Clustering in Process Mining. *Proc. 6th. Int. Conference on Business Process Management*, 109–120, Milan, Italy, 2008.

Part II

Poster Presentations

Human prophylaxis driven by risk may cause oscillations in sexually transmitted diseases

Benjamin Steinegger [*]

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
`benjaminfranzjosef.steinegger@urv.cat`

Several infectious diseases display oscillations in the incidence through time. In a variety of cases, the subsequent outbreaks are caused by seasonal, exogenous events, such as the increase of influenza cases in winter, or the increase of vector-borne diseases during rainy seasons. However, there are diseases like mycoplasma pneumoniae and syphilis which display non-seasonal periodic oscillations with a period of 3-7 years and 8-11 years [1], respectively.

Different mathematical models aim to capture these oscillations, either by explicitly including a period in the transmission rate, by allowing link rewiring in contact networks, or by considering models with temporary immunization. The aim of these models is to incorporate the behavioral response of individuals, which eventually leads to sustained oscillations in the disease incidence.

In this work, we present a stochastic, yet analytically tractable, epidemic spreading model coupled with a two-strategy evolutionary game, which reflects the individuals decision on whether to take preventive measures. Agents have information about the global extent of the disease, which serves as an assessment of their infection risk. The dependence of the evolution of the disease incidence on the individual decision on prophylaxis, is sufficient for the emergence of sustained oscillations as presented in Fig.1.

We describe the disease by an SIS compartmental model, as it is the case for many sexually transmitted diseases. We model the disease spreading on synthetic, heterogeneous networks, which mimic the characteristics of real world sexual contact networks. Finally, we propose plausible and efficient mechanisms to damp the oscillations. We show that targeted interventions, which are triggered as the disease incidence starts increasing, are much more effective than constant interventions of the same amplitude. In this sense, our study adds to the design of prevention campaigns, which do not only focus on perceived but real risks, in order to ameliorate human prophylactic behavior and contain future outbreaks of sexually transmitted diseases.
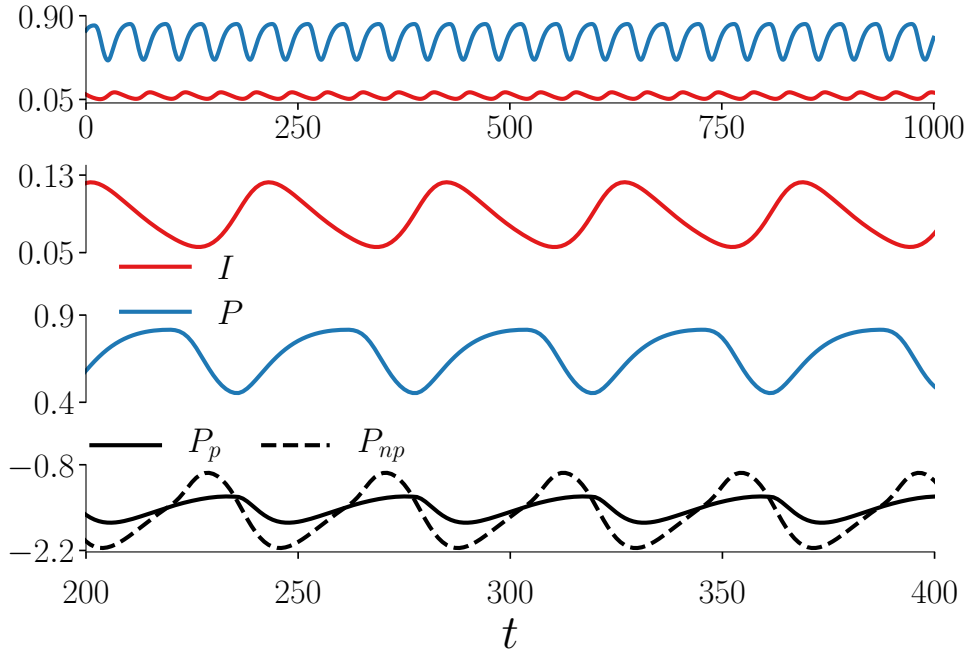
---

[*] PhD advisor: Alex Arenas & Clara Granell

Fig. 1. Fraction of infected, I, and protected, P, agents as a function of time, t, in red and blue, respectively. The last panel shows the payoffs for protected and not protected agents. We see that as the incidence is increasing, protection becomes more beneficial and individuals adopt protection until the disease is under control again. However, the protection uptake leads to a decrease in the incidence making protection less beneficial and individuals stop protecting causing a new cycle.

References

[1] N.C. Grassly et al. Host immunity and synchronized epidemics of syphilis across the United States. *Nature.*, 433:417 EP, 2005.

Registering a single 2D image to a 3D geometric model: from Traditional Machine Learning to Deep Learning

Saddam Abdulwahab *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
saddam.abdulwahab@urv.cat

1 Indications

In computer vision research domain understanding the scene and objects is an important task. Where used in various object registration tasks and different computer vision applications such as human pose estimation, face identification, and robotics. Therefore, finding a relationship between image pixels and 3D geometric features is a critical issue for image scene understanding. Viewpoint estimation, recognize relations between 2D images and 3D models and reconstructing depth image from a single RGB image is an essential component in understanding the 3D geometry of a scene. However, the depth map and viewpoint prediction from a single 2D image are still major challenge [1]. Since depth images represent only objects' shape compared to intensity image, whereas the intensity image is relative to the view-points, texture and lighting conditions. Thus, it is essential to extract the 2D and 3D features of the two modalities to match them to find the correct view. But the matching of a 2D image to a 3D model is considered a difficult task since the appearance of an object dramatically depends on its intrinsic characteristics (e.g., texture and color/albedo), and extrinsic characteristics related to the acquisition (e.g., the camera pose and the lighting conditions). The 2D/3D matching problem is mainly about knowing the appropriate representation method that can be used for extracting features in both 2D and 3D data and how to match entities between the two modalities in a common representation.

The existing systems [2,3,4] of viewpoint estimation from RGB image proposed image-to-model registration to estimate the 3D pose of the object use machine learning techniques. In addition, there are several methods based on deep learning techniques. For example, have been proposed for 3D shape generation from a single color image. For instance, there are a concurrent work of [5], which aims to solve the problem of estimating a depth map from a single

---

* PhD advisor: Hatem A. Rashwan and Domenec Puig

image of a face. Based on the employed approach we categorize the systems into two categories: machine learning models and deep learning models. The next subsections explain briefly those systems.

2 Machine Learning Models for 2D/3D Registration

Machine learning (ML), such as neural network (NN) and support vector machine (SVM) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead. Also, machine learning based models can extract patterns from data, there is one main limitation is that they are highly dependent on hand-crafted features which is time-consuming. For the problem of automatically aligning 2D intensity images with a 3D model has been recently investigated in depth. In the general case, the proposed solution will be image-to-model registration to estimate the 3D pose of the object.

For various registration methods, the 3D models have been represented in different ways (e.g., depth or synthetic images) and then the features extracted from the query and rendered images are matched. In [6,7], correspondences were obtained by matching SIFT feature descriptors between SIFT points extracted from the color images and from the 3D models. However, establishing reliable correspondences may be difficult due to the fact that the features in 2D and 3D are not always similar, in particular, because of the variability of the illumination conditions during the 2D and 3D acquisitions. Other methods relying on higher level features, such as lines [8], planes [9], building bounding boxes [10] and Skyline-based methods [11] have been generally suitable for Manhattan World scenes and hence applicable only in such environments.

Recently, the histogram of gradients, HOG, detector [12,13] or its fast version proposed [14] have been also used to extract the features from rendering views and real images. These approaches have not evaluated the repeatability between the correspondences detected in an intensity image and those detected in rendered images. In turn, 3D corner points have been detected in [15] using the 3D Harris detector and the rendering Average shading gradients (ASG) images have been generated for each detected point. For a query image, similarly, 2D corner pixels are detected in multi-scale. Then, the gradients computed for patches around each pixel are matched with the database containing ASG images using HOG descriptor. This method still relies on extracting gradients of intensity images affected by textures and background yielding erroneous correspondences.

In our work based on machine learning techniques, we used the concept of curvilinear saliency, related to curvature estimation, for extracting the shape information of 2D and 3D modalities. we proposed to cluster the depth images into groups based on Clustering Rule-based Algorithm (CRA). In order to

reduce the matching space between the intensity and depth images, a 2D/3D registration framework based on multi-class Support Vector Machine (SVM) is then used. SVM predicts the closest class (i.e., a set of depth images) to the input image. Finally, the closest view is refined and verified by using RANSAC. The effectiveness of the proposed registration approach has been evaluated by using the public PASCAL3D+ dataset. Figure 1 shows the Machine Learning Models for 2D/3D registration algorithm.
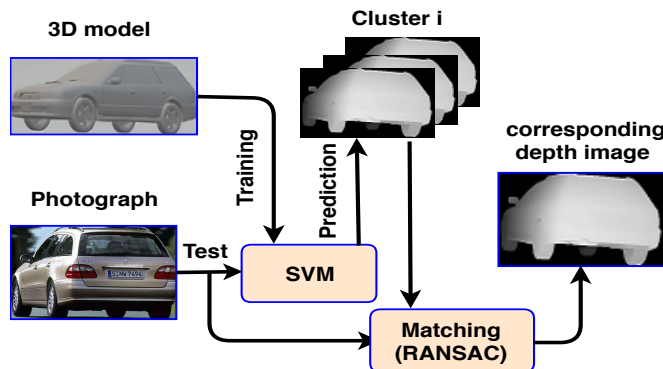

Fig. 1. Machine Learning Models for 2D/3D registration algorithm.

3 Deep Learning Models for 2D/3D Registration

Deep learning is a part of machine learning that has revolutionized the area of artificial intelligence. It is widely used in computer vision and natural language processing, yielding best outcome and outperforming most of the state-of-the-art approaches. In viewpoint estimation and depth prediction from an RGB image, there are most of the state-of-the-art methods based on deep learning. They provide automatic feature extraction and both richer representation capabilities and better performance than traditional hand-crafted feature-based techniques. The researchers have leveraged Recurrent Neural Networks (CNNs) and Generative Adversarial Networks (GANs) to solve the targeted viewpoint estimation and depth predict a problem.

Nowadays, with the significant progress with developing deep learning models, several approaches based on deep learning models have been proposed to predict depth maps from a single image. The authors of [16] proposed a method for 3D object detection and pose estimation from a single image using a deep CNN. To estimate an objects full 3D pose and dimensions from a 2D bounding box, they used a discrete-continuous CNN architecture network for orientation prediction and a practical choice of box dimensions as regression parameters, to estimates stable and accurate posed 3D bounding boxes

without additional 3D shape models or sampling strategies with complex pre-processing pipelines. However, the main problem during the training on real images to predict the poses in a very constrained subspace. In turn, [17] developed a CNN method using deep learning and geometry. This approach consists of two networks the one presents a 3D pose estimation approach for object categories. Moreover, the other one uses render depth images from 3D models under the estimated pose as a before retrieving 3D models and extract image descriptors from the real RGB image and the synthetic depth images. In this way, they match the computed descriptors to retrieve the closest 3D model and to try to retrieve 3D models from ShapeNet, which accurately represent the geometry of objects in RGB images.

In our work based on deep learning techniques, we used big progress in deep learning techniques to solve the problem of estimating depth from a single image. Besides, we used the generated depth image to predicting the 3D pose of an object in an image. Our model consists of two successive networks. The first network is based on a Generative Adversarial Neural network (GAN) for estimating a depth map from the input image. Figure 2 shows the Deep Learning Models for Depth Prediction and Viewpoint Estimation. A CNN for a regression task is then used to predict the 3D pose from the generated depth. However, a hard issue makes difficulties to estimate the depth and the 3D pose using deep models is the lack of training data with depth and view annotations. Thus, this work assumes a cross-domain training procedure to train the proposed model. We use 3D CAD models corresponding to objects appearing in real images to rendering depth images of different viewpoints and then using these rendering images as a guide for the GAN network to learn how to convert from the image domain to depth domain. The proposed model is evaluated on the PASCAL 3D+ dataset obtaining results outperforming the state-of-the-art models.
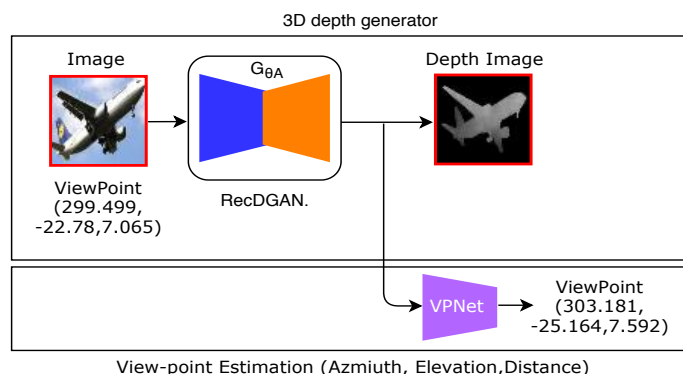

Fig. 2. Deep Learning Models for Depth Prediction and Viewpoint Estimation.

4 Conclusions and future work

In this work, we presented automatic 2D/3D registration approach compensating the disadvantages of rendering a large number of images of 3D models (i.e., depth images) by reducing the matching space between the 2D intensity and 3D depth images based on machine learning techniques. And we proposed a novel cross-domain deep model for estimating the depth and the 3D pose of an object in a color image comprising two successive networks: Generative Adversarial Network (RecDGAN) for predicting the depth images with a regression CNN network for estimating the viewpoint (VPnet) based on deep learning techniques. Future works aim to use this system to use the proposed model in an object grasping framework based on a single RGB camera.

References

[1] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-d objects from appearance," *International journal of computer vision*, vol. 14, no. 1, pp. 5–24, 1995.

[2] H. A. Rashwan, S. Chambon, P. Gurdjos, G. Morin, and V. Charvillat, "Towards multi-scale feature detection repeatable over intensity and depth images," in *Image Processing (ICIP), 2016 IEEE International Conference on*, pp. 36–40, IEEE, 2016.

[3] H. A. Rashwan, S. Chambon, P. Gurdjos, G. Morin, and V. Charvillat, "Using curvilinear features in focus for registering a single image to a 3d object supplemental materials," *IEEE Transactions on Image Processing*, pp. 1–1, 2019.

[4] S. Abdulwahab, H. A. Rashwan, J. Christino, S. Chambon, and D. Puig, "Effective 2d/3d registration using curvilinear saliency features and multi-class svm,"

[5] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8789–8797, 2018.

[6] T. Sattler, B. Leibe, and L. Kobbelt, "Fast image-based localization using direct 2d-to-3d matching," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 667–674, IEEE, 2011.

[7] Y. Y. Lee, M. K. Park, J. D. Yoo, and K. H. Lee, "Multi-scale feature matching between 2d image and 3d model," in *SIGGRAPH Asia 2013 Posters*, p. 14, ACM, 2013.

[8] C. Xu, L. Zhang, L. Cheng, and R. Koch, "Pose estimation from line correspondences: A complete analysis and a series of solutions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1209–1222, 2017.

[9] M. Tamaazousti, V. Gay-Bellile, S. N. Collette, S. Bourgeois, and M. Dhome, "Nonlinear refinement of structure from motion reconstruction by taking advantage of a partial knowledge of the environment," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 3073–3080, IEEE, 2011.

[10] L. Liu and I. Stamos, "Automatic 3d to 2d registration for the photorealistic rendering of urban scenes," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 137–143, IEEE, 2005.

[11] S. Ramalingam, S. Bouaziz, P. Sturm, and M. Brand, "Geolocalization using skylines from omni-images," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pp. 23–30, IEEE, 2009.

[12] M. Aubry, D. Maturana, A. A. Efros, B. C. Russell, and J. Sivic, "Seeing 3d chairs: exemplar part-based 2d-3d alignment using a large dataset of cad models," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3762–3769, 2014.

[13] J. J. Lim, A. Khosla, and A. Torralba, "Fpm: Fine pose parts-based model with 3d cad models," in *European Conference on Computer Vision*, pp. 478–493, Springer, 2014.

[14] C. B. Choy, M. Stark, S. Corbett-Davies, and S. Savarese, "Enriching object detection with 2d-3d registration and continuous viewpoint estimation," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pp. 2512–2520, IEEE, 2015.

[15] T. Plötz and S. Roth, "Automatic registration of images to untextured geometry using average shading gradients," *International Journal of Computer Vision*, vol. 125, no. 1-3, pp. 65–81, 2017.

[16] A. Mousavian, D. Anguelov, J. Flynn, and J. Košecká, "3d bounding box estimation using deep learning and geometry," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pp. 5632–5640, IEEE, 2017.

[17] A. Grabner, P. M. Roth, and V. Lepetit, "3d pose estimation and 3d model retrieval for objects in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3022–3031, 2018.

Benefits of the Statistical and Machine Learning Models in Games Designed for Children with Cerebral Palsy

Saddam Abdulwahab *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
`saddam.abdulwahab@urv.cat`

1 Indications

Cerebral palsy (CP) is a group of permanent movement disorders that appear in early childhood and one of the most frequent causes of disability in childhood, with an incidence of 2 per 1000 live births. In the EU, there are 1.3 million out of 15 million people with CP in the world. symptoms include poor coordination, stiff muscles, weak muscles, and tremors. There may be problems with sensation, vision, hearing, swallowing, and speaking. Where we can say this neurological disorder affects body movement, balance and posture. Other symptoms include seizures and problems with thinking or reasoning, which each occur in about one-third of people with CP.

Children with a disability enjoy many of the same hobbies and activities like their typically developing peers. Even though Computer gaming has come a long way since the advent of the ATARI or Commodore 64, with the EU gaming industry worth an estimated 180.1 billion per year. However, children with motor impairments encounter some of the obstacles when attempting to play commercial gaming systems, compared with their peers. Because the gameplay itself can require complex control from any player but for children with CP is very difficult. In this context [1], video games, especially exergames, are a very promising way to enable young people with cp to perform the exercise they need to break the circle of conciliation while allowing them to communicate with others in ways that are enjoyable from the comfort of home. Exergames are a combination of exercises (or effort) and video games. The reviews of these type of Exergames indicate that they have positive effects on both motivations for effective participation in rehabilitation and vulnerable jobs. In this paper, we will highlight what the benefits form the statistical and machine learning models in the games for the children with cp In GABLE

Project. Figure 1 shows the games for the children with cerebral palsy In GABLE Project.

Fig. 1. Games for the Children with Cerebral Palsy In GABLE Project.

2 Proposed Approach

One of the important aspects of the GABLE project is to gather useful information from initial data to help caregivers in tracking patients 'cases and to know how well the situation improves when the patients. Also, it provides the caregiver's recommendations on changing the difficulty level of a certain game and recommends more challenging games for the patients when he decides to play. In this section, we will explain the main steps of the approach. *Statistical Analysis and Prediction and Recommendation in Machine Learning Models.*

2.1 Statistical Analysis

In this subsection, we provided basic analysis to the caregivers and parents with the website platform [2], which allows parents and caregivers to track the progress of the patients using the score of the games. The module analyzed all data collected and displayed the statistical data on the types of games a patient has played, the progress she/he has made while playing a certain game repeatedly, etc. and gives an overall view of all the activity of the player, including game time, preference to a certain type of game, and assist the caregiver in the active monitoring of a patient. which help them to see how well

the children with the disease improve their performance. using this statistical we help the caregivers can predict the score of their patients in order to see if playing a particular game will be useful for their patients or not. Figure 2 shows a general overview of the statistical analysis platform.
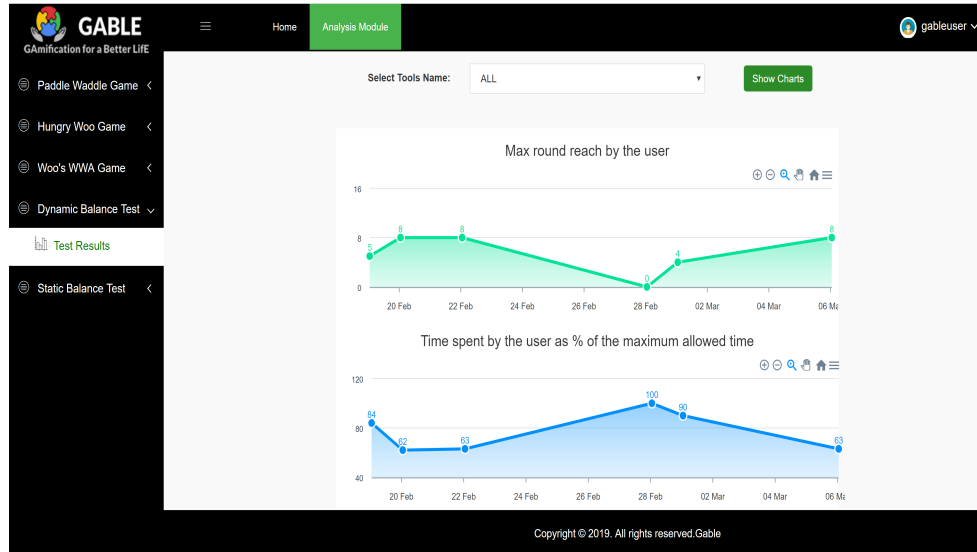
Fig. 2. General overview of the Statistical Analysis Platform.

2.2 Prediction and Recondition System

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead. Also, machine learning based models can extract patterns from data. There is one main limitation is that they are highly dependent on handcrafted features which is time-consuming. In this subsection, we used the data gathered on the games played by the patients and their progress within a game to designing a machine learning module to provide the caregiver's recommendations on changing the difficulty level of a certain game. And as a measure of how many patients are playing any game, or how high it was scored by the patients and other caregivers and parents. Using this rating system, the module can then recommend a certain game for a patient, or the caregiver/parent. Figure 3 shows a general overview of the Prediction Model.
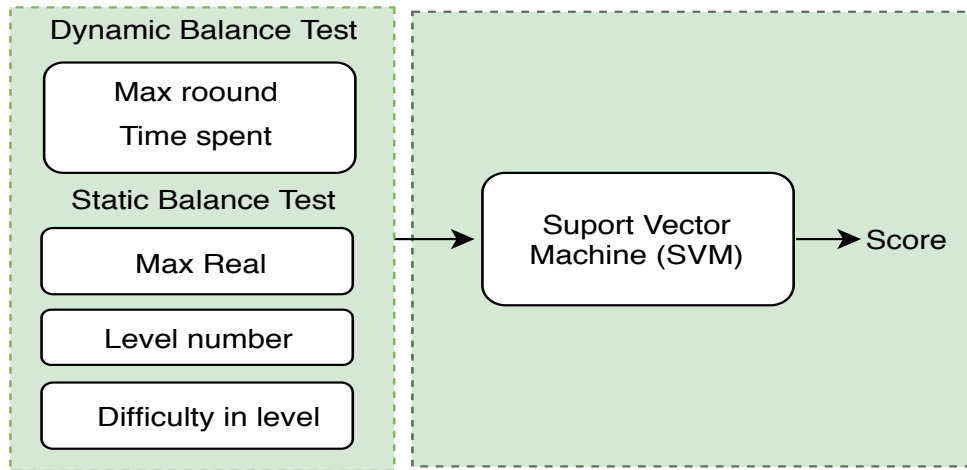
Fig. 3. General overview of the Prediction Model.

3 Conclusion and Future work

In this work, we designed a website platform to analysis patients data for the caregivers and developed a machine learning module to provide the caregiver's recommendations on changing the difficulty level of a certain game. In the Future work, we will integrate that model into the development platform and game-authoring tools, which mainly performs three different tasks including statistical analysis, prediction analysis, and recommendation system.

References

[1] "gable app." https://gable.app. last accessed: 01-April-2019.

[2] "statistical model." https://projectgable.eu/analytics. last accessed: 10-April-2019.

A distributed Incident alert priority-based dissemination protocol for VANETs

Assia NAJA *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
`naja.assia@gmail.com`

1 Abstract

This work presents a new rebroadcasting algorithm called Dynamic Hybrid Broadcasting Protocol *DHBP*, designed to enhance the performance of the counter-based protocol. Then we examine four rebroadcasting protocols with different parameters: *counter-based, probabilistic-based, flooding* and *HCAB* (*hop count-aided broadcasting*). Performance is assessed with the NS2 network simulator, using a realistic mobility model and different traffic parameters. For the simulations, we considered three parameters: saved rebroadcast, latency and reachability. We then analyzed the performance with three scenarios, in which we varied the speed of nodes, the number of nodes, and the area size.

2 Introduction

The increase in the number of vehicles all around the world induced an increase in the number of vehicle accidents and fatalities with more than 1.2 million victims annually across the globe [1]. Statistics relate that vehicle accidents are the primary cause of death for humans in both Europe and the USA [2].

In Morocco for example, the vulnerable road users (pedestrians and users of two-wheelers) are the first victims of road accidents, which remain the most concerned recording 52.96% of deaths category, followed by car users who present 36.36% of all killed pedestrians [3]. The related dangers are considered to be a solemn problem that society nowadays is facing. Thus, the VANETs become the mobile Ad-hoc networks class with most challenge to have a major impact on enhancing road safety, traffic efficiency, comfort to passengers, and traffic management through a multitude of new pervasive applications developed in this context [4].

\- Inform users in real-time about road conditions can help them to better anticipate some dangers.

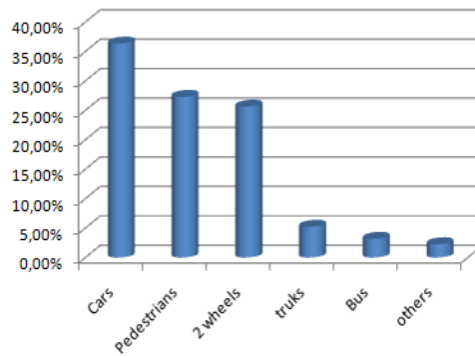* PhD advisor: Pr. Domenec Puig

Fig. 1: Distribution of deaths by category of users and structure of the vehicles involved.
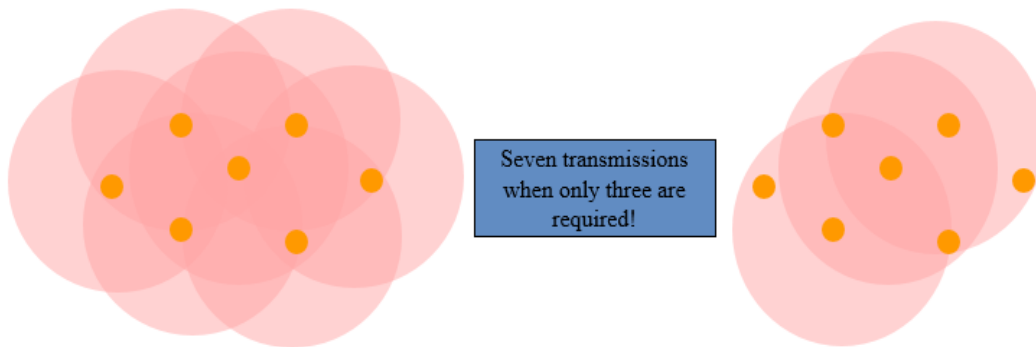

Fig. 2: Storm problem

Figure. 2 presents a situation of flooding, which means that each node that receives an information will rebroadcast it automatically.This leads to to the broadcast storm problem.

2.1 Broadcast storm problem

- Caused by the flooding broadcasting
- Redundant rebroadcasts (Resulting a huge number of redundant packets)
- Heavy contention could exist Collisions are more likely to occur

2.2 Research question

- How to derive an efficient scheme for broadcasting in a VANET?
- How to maximize the amount of disseminated data while minimizing the number of redundant messages and keeping tradeoff between lower latency and good reachability.

3 Methodology

$$DMF = \begin{cases} P_{r_{V_i}}(t) & c = 1 \ \& \ d_i < D \\ P_{r_{V_i}}(t) * (1 - \frac{d_i}{D}) * (1 - \frac{c}{C}) & 2 \leq c \leq C \ \& \ d_i < D \\ 0 & c > C \ \& \ d_i > D \end{cases} \quad (1)$$

$$P_{r_{V_i}}(t) = \begin{cases} P_{r_0} \cdot \left(1 - \tanh(\frac{t - t_0}{0.5T})\right) & t_0 \leq t < T \\ 0 & t \geq T \end{cases} \quad (2)$$

$$\xi = \begin{cases} tanh(\frac{s}{S}) & s < S \\ \infty & s > S \end{cases} \quad (3)$$

4 Simulation Results

Simulation Parameter	Value
Scenarios used	800 m x 800 m Grid
	Highway
	Real Map of Agdal (Rabat, Morocco)
	of 2500 m * 1250 m
Transmission range	250 m
Maximum Speed of Nodes	
Grid Map	20, 60, 85(km/h)
Highway	120(km/h)
Real Map of Agdal	60(km/h)
Bandwidth	2 Mbps
Broadcast Sources	10
Simulated of nodes	50, 100, 150
Packet size(bytes)	1000
Simulation time(Secs)	1000
Trials	20

Table 1: ***Simulation Parameters***

- The delay decreases as the number of nodes increases for all speeds.
- Counter- and probability-based protocols have a higher average delay if compared with DHBP.
- SRB increases as we increase the number of nodes for all speeds.
- DHBP has relatively better SRB in comparison to other algorithms under different node densities and speeds.
- Counter-based scheme (especially when choosing a threshold of 3) had significant saved rebroadcast ratio but suffer from a high delay.

Algorithm 1 *Pseudo code of the proposed DMF (Decision Making Function)*

1: **Input**: Received packet by broadcast
2: **function** DECISION(Packet)
3:     **Step 1:** Upon receiving a packet for the first time
4:     **if** The packet is received before **then**
5:         Discard the packet and exit;
6:     **else**
7:         **Step 2:**
8:         Initialize a counter c=1;
9:         DMF=1;
10:        Compute d_1;
11:        **Step 3:**
12:        Generate an assessment time ξ;
13:        **Step 4**: Wait until the generated time ξ expires
14:        **if** the same packet is received again **then**
15:            Increase the counter by 1;
16:        **end if**
17:        **Step 5:** When ξ expired
18:        Compute d_2;
19:        **if** $(d_2\text{-}d_1) >0$ && $d_2 >$D **then**
20:                                      ▷ (D is a specified threshold)
21:            Discard the packet and exit;
22:        **else**
23:            **if** c==1 **then**
24:               DMF=1;
25:               Rebroadcast the packet and exit;
26:            **else if** c=<C && c>=2 **then**
27:               Compute DMF;
28:               Nr=rand(0,1);
29:               **if** (Nr<=DMF) **then**
30:                  Rebroadcast the packet and exit;
31:               **else**
32:                  Discard the packet and exit;
33:               **end if**
34:            **else if** c>C **then**
35:               DMF=0;
36:               Discard the packet and exit;
37:            **end if**
38:        **end if**
39:     **end if**
40:     **Return** Retransmitting or discarding the received packet
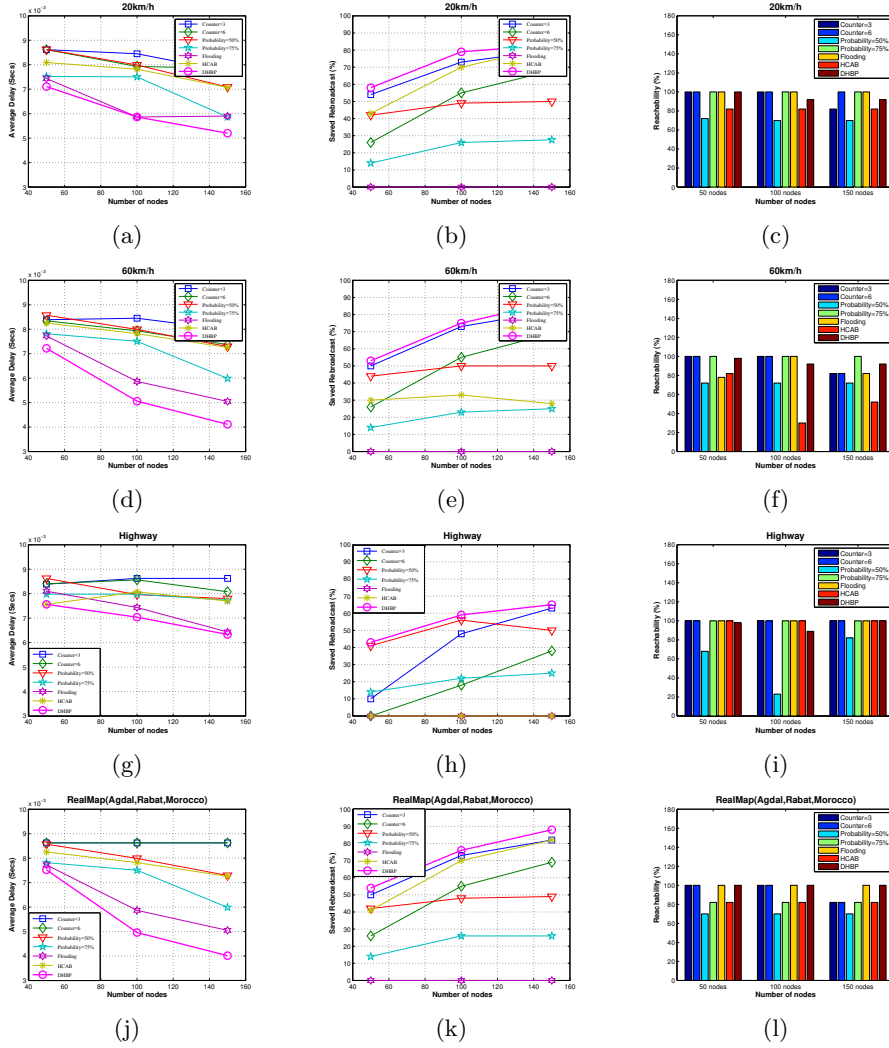41: **end function**

Fig. 3: Comparison of flooding, HCAB, DHBP, counter and probability-based protocols in term of average delay, SRB and reachability vs. number for 20 km/h and 60 km/h in a grid scenario and in highway and realmap scenarios

- The probability based scheme presents a moderate value of SRB (worse than those of the counter-based protocols) but suffer from the same problem which is the higher delay.
- DHBP offer the best compromise between higher SRB and lower delay.
- For all counter-based schemes, the reachability is 100%. However, the probability-based schemes have poor reachability due to the use of a small probability in sparse areas.
- DHBP shows almost relatively lower delay compared to other schemes for both highway and real map cases, higher SRB and higher reachability ratio.

5 CONCLUSION

- Problem: ailoring a rebroadcasting protocol for VANET is a challenging task
- Solutions:
  - We evaluated existing protocols using NS2 in several scenarios that varied node density and speeds
  - A new broadcast protocol is developed which copes with all requirements
- Results:
  - Generates fewer rebroadcasts
  - Decreases the average delay
  - Increases the reachability

References

[1] Road Safety Annual Report, 2017

[2] H. Hartenstein and K. Laberteaux, VANET: vehicular applications and internetworking technologies. John Wiley & Sons, 2009, vol. 1.

[3] Moroccan Commission Report Road accident statistics in Europe, 2012.

[4] P. A. Abdulla et C. Delporte-Gallet, Networked Systems: 4th International Conference, NETYS 2016, Marrakech, Morocco, May 18-20, 2016, Revised Selected Papers. Springer, 2016.

Anonymization of Textual Documents using Word Embeddings

Fadi Hassan [*]

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
`fadi.hassan@urv.cat`

1 Introduction

Automatic textual data anonymization consists of two main parts: detecting sensitive pieces of text that can be used to re-identify the entity to be protected, and protecting these sensitive pieces by removing or masking them.

The detection of sensitive entities can be approached in different ways. The simplest approach is based on Named-Entity Recognition (NER) and uses classifiers trained on large amounts of manually tagged data, which are able to detect a fixed sets of sensitive entities such as names, locations, dates, etc. However, this approach is severely limited because not all the sensitive information appearing in text fits in one of these pre-defined types and not all appearances of a certain entity type may disclose information on the entity to be protected. More sophisticated approaches are based on calculating the semantic relatedness between the linguistic entities in the text and the entity to be protected [3]. Semantic relatedness is assessed using distributional [5] or probabilistic models [1], which are built by calculating statistics on the (co-)occurrence of words. A recent trend in computational linguistics to measure the semantic relatedness between words is to use neural network-based word embedding models [6], which will exploit in our work.

Protecting sensitive entities, on the other hand, can be solved in similar way as anonymization of structured data. Once we detect the sensitive entities in a document, we can treat them as standard attributes. In anonymization of structured data bases attributes are categorized according to the re-identification risk: (i) Identifiers are attributes whose values are enough to re-identify individuals and the usual approach to data protection is to remove them. (ii) Quasi-identifiers are attributes that separately do not allow re-identification but whose combination may; the usual approach to data protection is to mask them by perturbing, generalizing or even removing them.

---

[*] PhD advisor: Jordi Soria-Comas and Josep Domingo-Ferrer, Additional author: David Sánchez

The next section explain briefly our method to tackle the anonymization of textual data.

2 The proposed method

To automatically detect the sensitive entities in text we characterize the linguistic terms appearing in a document according to the amount of information they reveal on the entity to be protected. We tackle this problem by means of word embedding.

Word embedding maps words or phrases from a vocabulary to vectors of real numbers. For these vectors of the words to be useful, we want similar words to be mapped to similar vectors. There are several ways to do this mapping. The current state of the art is based on 2-layer neural networks [6].

Our approach consists of three phases. First, we us a large corpus to train a word embedding model which capture the semantic relationships between all the terms appearing in a collection of documents. Second, we use the trained model to detect the terms in the document which have semantic relationship with the entity to be protected. Finally, we protect these sensitive terms. We explain these phases briefly in following subsections.

2.1 Training the model

The first phase of our method consists of the following steps:

- Data collection.
- Data pre-processing.
- Configuration of training parameters.
- Model training.

To obtain a word embedding model that accurately characterizes the relationship between words, we need a collection of documents describing the entities those we want to protect.

Secondly, since semantic inferences occur at a conceptual level, and concepts and entities are referred to in a text via noun phrases rather than individual words, we need to get vector representation of the concepts. Therefore, in the pre-processing step, we extract noun phrases (rather than individual words) and feed them as training data to the word embedding model.

Finally, we configure the training parameters so that the model fits with the purpose of detecting sensitive entities. In our work we use the standard implementation of neural network-based word embeddings (Word2Vec) [7]. This model has several parameters, which we adapt to our data protection goal as follows:

- Architecture: the skip-gram rather than bag-of-words because it is better for infrequent words.

- Window size: is set to the size of sentences, which usually 10.
- Dimension of the word vectors: in semantic similarity assessments a value of 300 gives good results.
- minimum number of appearances: it should be zero in the context of document anonymization because rare words (such as names or concrete addresses, which would appear once) are usually those that cause a greater risk because they usually refer to very specific (quasi-)identifying information [2].

The result of this phase are a set of word embedding vectors for all the noun phrases appearing in the collection of documents.

2.2 Detecting Sensitive terms

Once we get the trained model from the previous phase, we can use it to obtain a vector representation of each term in our input collection of documents. Because terms closely related in semantics tend to appear in similar contexts, this yields similar vectors [4]. We use these vector representations to evaluate the disclosure risk of each terms in the document by measuring the cosine-similarity between their vectors and the vector of the entity to be protected. Terms that have a similarity above fixed threshold t are considered sensitive.

2.3 Protecting Sensitive terms

Different strategies may be employed to protect sensitive terms in text as we mentioned before. So far, we just remove them, but we plan to use generalization for masking to get more utility from the anonymized document.

3 Conclusion and Future work

In this work, we presented a new method to detect sensitive terms in documents by using word embedding models. Our method is more general and less constrained that current methods based on NER. As future work, we can devise several ways to improve the results. First, we can improve the pre-processing step by incorporating morphological analyses by which derivative forms of the same words/phrases (e.g., singular/plural) can be identified and considered in aggregate. Also, we plan to design sanitization algorithms for the second part (protecting sensitive entities), using generalization rather than removal. For this, we plan to use large knowledge bases (e.g., YAGO or WordNet), whose concepts will also be incorporated into the word embedding model.

References

[1] D. Sánchez, M. Batet, A. Valls, and K. Gibert, "Ontology-driven web-based semantic similarity," *Journal of Intelligent Information Systems*, vol. 35, no. 3, pp. 383–413, 2010.

[2] D. Sanchez, M. Batet, and A. Viejo, "Automatic general-purpose sanitization of textual documents," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 853–862, 2013.

[3] D. Sánchez and M. Batet, "C-sanitized: A privacy model for document redaction and sanitization," *Journal of the Association for Information Science and Technology*, vol. 67, no. 1, pp. 148–163, 2016.

[4] M. Sahlgren, "The distributional hypothesis," *Italian Journal of Disability Studies*, vol. 20, pp. 33–53, 2008.

[5] S. M. Mohammad and G. Hirst, "Distributional measures of semantic distance: A survey," *arXiv preprint arXiv:1203.1858*, 2012.

[6] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, pp. 3111–3119, 2013.

[7] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

A study about the mathematical knowledge of pre-service teachers

Jaime Segarra Escandon [*]

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
`jaimerodrigo.segarra@urv.cat`

1 Introduction

The results obtained in the last international tests TIMSS (2015) (Third International Mathematics and Science Study) and PISA (2015) (Programme for International Student Assessment) evidence the weak mathematical knowledge of Primary and Secondary students, in most of the countries that apply those tests. These international evaluations contribute to improve the comprehension of the teaching and knowledge [2].

The TIMSS project quantifies the mathematics and science knowledge of the students, together with the context it occurs [9]. Specifically, TIMSS evaluates the mathematical knowledge in the four content domains: Numbers, Algebra, Geometry, and Data and Probability. On the other hand, it evaluates the three cognitive domains: knowing, applying and reasoning. In TIMSS 2011, Spain achieve 482 points (recall that the OCDE mean was 522). In TIMSS 2015, Spain scored 505 points (the OCDE mean was 525).

PISA evaluates 15 years students and analyze the knowledge of reading, mathematics and science. Several works in the literature show a clear concern due to the poor results obtained in PISA 2012 and PISA 2015 (e.g., [1], [7], [8]). Spain obtained 484 points in PISA 2012 in the mathematics test (notice that the OCDE mean was 494). In 2015, it obtained 486 points (the OCDE mean was 490).

Therefore, in the last evaluations of TIMSS and PISA (mathematics), Spain did not arrive to the OCDE mean. These poor results are related with the weakness that may appear during the pre-service teacher instruction, both in mathematical content and didactic knowledge [5]. Thus, it is crucial the study of the pre-service Primary teacher programs, since their knowledge influence the students knowledge. In fact, many investigations evaluate the knowledge of pre-service teachers (e.g., [4], [3], [6], [10], [11]).

---

[*] PhD advisor: Carme Julià

As professors in the Primary Education Bachelor Degree at Universitat Rovira i Virgili, we have detected some weakness that students presents in some mathematics contents and processes. This work aims at studying those weakness in order to improve the quality of the pre-service teacher program, focusing in the mathematics area.

Concretely, the goal of this research is to analyse the initial mathematical knowledge of the Primary Education Bachelor's students. The idea is to detect in which contents and processes they present more difficulties. The next step would be to design proposals to reinforce those concepts and processes.

2 Method

This study is designed for the first year students of the Primary Education Bachelor Degree at Universitat Rovira i Virgili in the 2018/19 academic year. A test was applied to the students that assist the day the classes started. Specifically, a total of 97 students did the test, which corresponds to the 71% of the students enrolled to the Bachelor. The 57% of students are women, while the 43% are men.

The instrument used to carry out the research consists of a collection of items of TIMSS 2011. The content domain of 2nd level of ESO was considered. Concretely, 20 items were selected: 12 items corresponding to the Numbers domain (P1-P12) and 8 items corresponding to the Geometry domain (P13-P20). All the items correspond to the cognitive domains of applying and reasoning. In order to make explicit the solving process that students follow, 11 items were reformulated. We will refer to these items as problems (P5, P6, P7, P8, P9, P10, P11, P12, P17, P18, P19 y P20). The other items maintain the TIMSS structure and we refer to these as objective items. The whole test is presented in the Appendix.

After applying the test to the students, it was qualified and a data base with all the results was generated. The problem items are qualified as 0 if the answer and process are wrong, 0.5 if the answer or the process are corrects and 1 if both are correct. The test is qualified in a scale from 0 up to 10. Then, the statistic analysis is carried out by using the R software.

The reliability study of the test gave a Cronbach Alfa of 0.74, which is acceptable.

3 Results

The analysis of the results is presented in two different sections. Namely, results according to the content domain and results according to correct, unanswered and wrong questions.

3.1 Results by content domain

Figure 1 show the mean obtained in the items corresponding to the content domain of Numbers. The mean corresponding to the 12 items of this content domain is 6.54 and the standard deviation is 1.65.
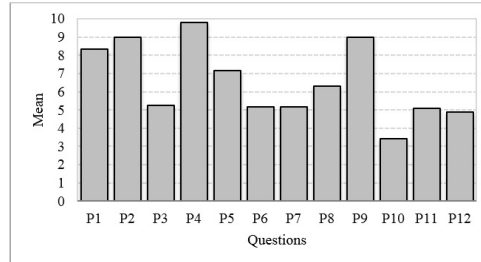

Fig. 1. Obtained mean in the items of the content domain of Numbers

It can be seen in Figure 1 that P4 is item with the highest score, with a mean of 9.79. This item belongs to the knowledge domain of Numbers and to the cognitive domain of application. It is related to the fractions and decimals. The item P10, on the contrary, is the item with the lowest score, 3.40. This item belongs to the cognitive domain of application and it is related to proportion and percentage.

Analogously, Figure 2 displays the mean obtained in each of the items in the content domain of Geometry. The obtained mean in this domain is 5.54 and the standard deviation 2.57.
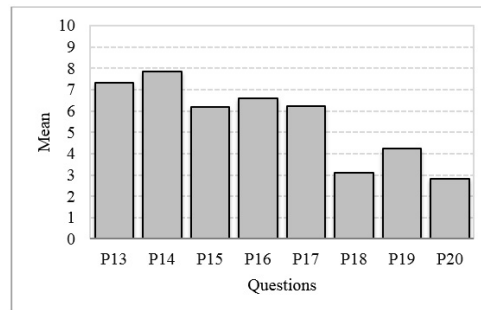
Fig. 2. Obtained mean in the items of the content domain of Geometry

Figure 2 shows that item P14 is the item with the highest mean, 7.84. Notice that the highest score is quite smaller than the one obtained in the domain of Numbers. Item P14 belongs to the cognitive domain of application and it is focussed on the geometric figures and spatial reasoning. On the other

hand, P20 is the item in which the students present more difficulties. The obtained mean is 2.84. This item belongs to the cognitive domain of reasoning and it is related to geometric measures.

Figure 3 displays all the scores obtained in each of the content domains (Numbers and Geometry). A boxplot is used in order to show the quartiles of the distribution. Recall that in the Numbers domain, the 50% of students obtain scores below 6.67. In the Geometry domain, the median is 5.63, which means that the 50% of students obtain scores below that value.
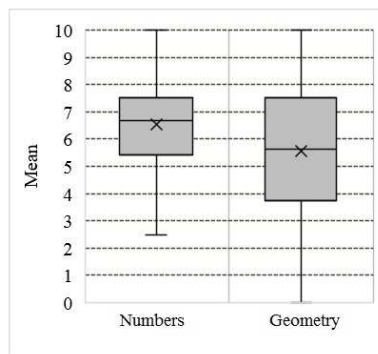
Fig. 3. Scorer obtained in each content domains (Numbers and Geometry)

Additionally, a statistic test has been carried out in order to study if the difference between the mean obtained in each content domain is statistically significant. Specifically, a t-student test with $\alpha = 0.005$ was applied, given $\rho = 0.0015$. Hence, the difference between both groups is statistically significant.

3.2 Results according to correct, blank and wrong answers

This section analyse the answers the students give to the items referred to problems (P5, P6, P7, P8, P9, P10, P11, P12, P17, P18, P19 y P20). Figure 4 displays the percentage of students that responded correctly, left the answer blank or responded wrongly.

Figure 4 shows that P10 is the most difficult problem for the students in the Numbers content domain. Notice that only 29% of the students solve it correctly, the 53% of them make errors and 18% of them left the problem in blank. This item belongs to the cognitive domain of application. Problem P9, on the contrary, is the one of the best results: 88% of correct answers, 10% of wrong answers and only 2% in blank. This item belongs to the cognitive domain of knowledge.

In the content domain of Geometry, P20 gives the worst results: 19% of the students answer correctly, 42% wrongly and 39% left it in blank. This item corresponds to the cognitive domain of reasoning. The best results, on the
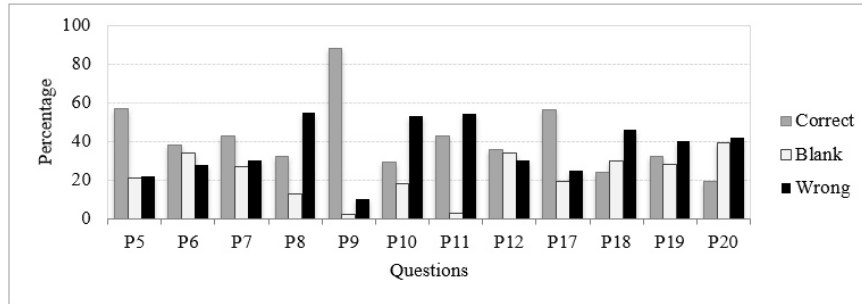
Fig. 4. Percentage of correct, blank and wrong answers, for every problem

contrary, are obtained in P17: 56% of correct answers, 25% of wrong answers and 19% in blank. This item belongs to the cognitive domain of knowledge.

Besides, it can be seen in Figure 4 the high percentage of students that do not write the process of the problems. In some cases, the percentage is higher than 28% (P6, P12, P18, P19 and P20). Other items present a percentage of error higher than 30% (P7, P8, P10, P11, P12, P18, P19 and P20). Finally, there are many items in which the students obtain low percentages of correct answers (below 40%): P6, P8, P10, P12, P18, P19 and P20.

4 Conclusions

This research studies the initial mathematical knowledge of the first year students of the Primary Education Bachelor Degree. Results evidence a statistically significant difference between scores obtained in the Numbers and Geometry domains, being higher the mean obtained in the Numbers domain. This work also analyses the answers that students give to the problems distinguishing among correct, blank and wrong answers.

The low percentage of students that give correct answers to the problems should be worrying. In the Numbers content domain, in 6 up to 8 items the percentage of correct answers is lower than 50%, while in the Geometry domain that percentage is not achieve in 4 up to 5 items.

On the other hand, results show that the items with highest percentage in blank, items with a mean below 5 and items with a high percentage of wrong answers belong to the cognitive domain of application or of reasoning.

This work propose to reinforce the observed weakness the students present during the Primary Education Bachelor Degree. Specifically, the teaching and learning classes of Mathematics should consider the difficulties the students find in Numbers and Geometry and also in the cognitive domain of reasoning. Additionally, the current results should be use to think about the teaching and learning mathematics at Primary and Secondary School, since the errors the students make correspond to concepts and processes taught in those levels.
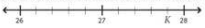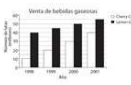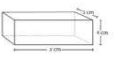
Appendix
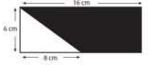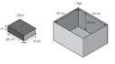
P1. ¿Which of these shows how 36 can be expressed as a product of prime factors? a. 6×6 b. 4×9 c. 4×3×3 d. 2×2×3×3	P2. ¿Which fraction is equivalent to 0,125? a. $\frac{125}{100}$ b. $\frac{125}{1.000}$ c. $\frac{125}{10.000}$ d. $\frac{125}{100.000}$
P3. ¿Which shows a correct method for finding $\frac{1}{3} - \frac{1}{4}$? a. $\frac{1-1}{4-3}$ b. $\frac{1}{4-3}$ c. $\frac{3-4}{3×4}$ d. $\frac{4-3}{4×3}$	P4. ¿Which number does K represent on this number line? 26 27 K 28 a. 27,4 b. 27,8 c. 27,9 d. 28,2
P5. ¿Which number is equal to $\frac{3}{5}$? a. 0.8 b. 0.6 c. 0.53 d. 0.35	P6. $\frac{4}{100} + \frac{3}{1000}$ a. 0.043 b. 0.1043 c. 0.403 d. 0.43
P7. The fractions $\frac{4}{14}$ y $\frac{\square}{21}$ are equivalent. a. 6 b. 7 c. 11 d. 14	P8. A workman cut off $\frac{1}{5}$ of a pipe. The piece he cut off was 3 meters long. ¿How many meters long was the original pipe? a. 8 b. 12 c. 15 d. 18
P9. 42,65÷5,748= Answer: _____	P10. Ana y Jenny divide 560 zeds between them. If Jenny gets $\frac{3}{8}$ of the money, ¿How many zeds will Ana get? Answer: _____
P11. Carla is packing eggs into boxes. Each box hold 6 eggs. She has 94 eggs. ¿What is the smallest number of boxed she needs to pack all the eggs? Answer: _____ box.	P12. The graph shows the sales of two types of soft drink over 4 years. If the sales trends continue for the next 10 years, determine the year in which the sales of Cherry Cola will be the same as the sales of Lemon Cola. a. 2003 b. 2004 c. 2005 d. 2006
P13. The large of side of each of the small square represents 1 cm. Draw an isosceles triangle with a base of 4 cm and a height of 5 cm.	P14. The figure below shows a shape made up of cubes that are all the same size. There is a hole all the way through the shape. ¿How many cubes would be needed to fill the hole? a. 6 b. 12 c. 15 d.18
P15. The volume of the rectangular box is 200 cm³. ¿What is the value of x? Answer: _____	P16. A piece of paper in the shape of a rectangle is folded in half as shown in the figure below. It is then cut along the dotted line, and the small piece that is cut is opened. What is the shape of the cutout figure? a. an isosceles triangle. b. two isosceles triangles. c. a right triangle. d. an equilateral triangle.
P17. The perimeter of a square 36 cm. ¿What is the area of this square? a. 81 cm² b. 36 cm² c. 24 cm² d. 18 cm²	P18. The area of a square is 144 cm². ¿What is the perimeter of the square? a. 12 cm b. 48 cm c. 288 cm d. 276 cm
P19. In the figure below, what is the area of the shaded region in cm²? a. 24 b. 44 c. 48 d. 72	P20. Ryan is packing books into a rectangular box. All the books are the same size. ¿What is the largest number of books that will fit inside the box? Answer: _____

Fig. 5. Test used in this research

References

[1] C. Berliner The many facets of PISA. *Teachers College Record.*, 117(1):1–20, 2015.

[2] R. Bybee TScientific literacy and contexts in PISA 2006 Science. *Journal of Research in Science Teaching.*, 46(8): 862–864, 2009.

[3] E. Castro. Resolución de Problemas: Ideas, tendencias e influencia. *En R. Luengo, B. Gómez, M. Camacho y L. Blanco (Coord.). Investigación en Educación Matemática XII, Badajoz: Sociedad Extremeña de Educación Matemática-SEIEM* , 2008.

[4] A. Flores, P. Alfaro, E. González, N. Rojas. El conocimiento especializado de un profesor de matemáticas: Un estudio de caso sobre la enseñanza de los conceptos básicos de función. *Uniciencia*, 32(1): 89–107, 2018.

[5] A. Gutiérrez, P. Gómez, L. Rico. Conocimiento matemático sobre números y operaciones de los estudiantes de Magisterio. *Educación XXI*, 19(1): 135–158, 2016.

[6] A. Huidobro, A. Méndez, L. Serrano. Del Bachillerato a la Universidad: las Matemáticas en las carreras de ciencias y tecnología. *Aula Abierta*, 38(1): 71–80, 2010.

[7] J. Hwang, K. Choi, Y. Bae, H. Shin Do Teachers' Instructional Practices Moderate Equity in Mathematical and Scientific Literacy?: an Investigation of the PISA 2012 and 2015. *International Journal of Science and Mathematics Education.*, 16(1): 25–45, 2018.

[8] Y. Lu, D. Bolt Examining the attitude-achievement paradox in PISA using a multilevel multidimensional IRT model for extreme response style. *Large-scale Assessments in Education.*, 3(1):1–18, 2015.

[9] M. Martín, R. Puertas. Comparativa de la eficiencia educativa de Europa y Asia: TIMMS 2015=Comparison of educational efficiency in Europe and Asia: TIMMS 2015. *Revista de educación*, (380):45–74, 2018.

[10] R. Nortes, A. Nortes. Competencia matemática, actitud y ansiedad hacia las Matemáticas en futuros maestros. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 20(3):145–160, 2017.

[11] R. Nortes, A. Nortes. Conocimientos matemáticos de futuros maestros en resolución de problemas de sexto de primaria. *Educatio Siglo XXI*, 36(3): 201–230, 2018.

This proceeding book contains the contributions presented at the 5th URV Doctoral workshop in Computer Science and Mathematics. The main aim of this workshop is to promote the dissemination of the ideas, methods and results that are developed by the students of our PhD program.

Departament d'Enginyeria

[DΣIM]

Informàtica i Matemàtiques
UNIVERSITAT ROVIRA I VIRGILI

ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA
Universitat Rovira i Virgili