Theoretical and Applied Informatics
Vol. 28 (2016), no. 1&2, pp. 1–12
DOI: 10.20904/281-2001

A New Proposal for Person Identification Based on the Dynamics of Typing: Preliminary Results

Krisztian Buza[1*] Dora Neubrandt[2†]

[1]Brain Imaging Center, Research Center for Natural Sciences
Hungarian Academy of Sciences, Budapest, Hungary

[2]Budapest University of Technology and Economics, Budapest, Hungary

Abstract The availability of cheap and widely applicable person identification techniques is essential due to a wide-spread usage of online services. The dynamics of typing is characteristic to particular users, and users are hardly able to mimic the dynamics of typing of others. State-of-the-art solutions for person identification from the dynamics of typing are based on machine learning. The presence of hubs, i.e., few instances that appear as nearest neighbours of surprisingly many other instances, have been observed in various domains recently and hubness-aware machine learning approaches have been shown to work well in those domains. However, hubness has not been studied in the context of person identification yet, and hubness-aware techniques have not been applied to this task. In this paper, we examine hubness in typing data and propose to use ECkNN, a recent hubness-aware regression technique together with dynamic time warping for person identification. We collected time-series data describing the dynamics of typing and used it to evaluate our approach. Experimental results show that hubness-aware techniques outperform state-of-the-art time-series classifiers.

Keywords person identification; dynamic time warping; hubness-aware machine learning

Received 17 FEB 2016 **Revised** 07 NOV 2016 **Accepted** 15 DEC 2016

1 Introduction

Conventional techniques for person identification range from passwords to biometric identification, such as fingerprints, iris-patterns, electroencephalograph-based and electrocardiograph-based person identification [1, 2, 3]. Online services, such as online banking or online courses, require cheap, widely accessible and reliable person identification techniques.

*E-mail: buza@biointelligence.hu
†E-mail: dori@biointelligence.hu

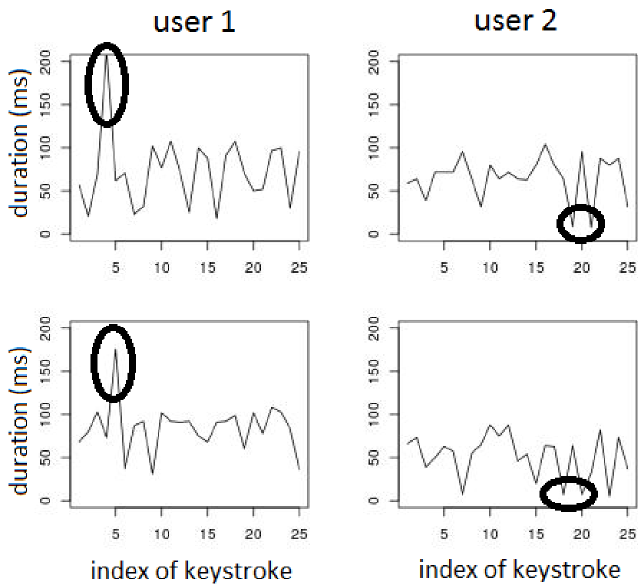

Figure 1 The duration of 25 consecutive keystrokes in case of two different users (in the left and right) when typing the same text.

It was shown that the dynamics of typing is characteristic to particular users [4], and most users are hardly able to mimic the typing dynamics of others [5]. It makes person identification based on typing patterns especially appealing in cases when the user is not necessarily interested to cooperate, such as testing the identity of students taking exams.

For example, in the case of an online course, users may be identified based on the dynamics of their typing, when they solve quizzes, tests, exams. Furthermore, user identification based on typing patterns could be potentially useful in case of regular on-campus courses and exams as well.

Moreover, in cases when high accuracy is required, such as online bank transactions involving significant amounts of money, person identification based on typing patterns could be used in combination with conventional identification techniques, such as passwords and authentication codes sent to the user in SMS or e-mail.

Although the dynamics of typing, e.g. the time series of the duration of keystrokes, is characteristic to users, it is evident that even the same user can not always type with the *exactly* same dynamics. This is illustrated in Fig. (1). The figure shows the durations of the first 25 keystrokes in the case of typing the same text by two different users. Each of the users typed the same text two times. The time series of *user1* are shown on the left of the figure, while the time series of *user2* are shown on the right of the figure. As one can see, the time series of the same user are more similar to each other than the time series of different users. In particular, a peak (i.e., an exceptionally long keystroke) close to the fifth position is characteristic to *user1*, whereas exceptionally short keystrokes close to position twenty are characteristic to *user2*. In case if we consider a large set of users, it may be challenging and time-consuming for human experts to identify patterns that can reliably distinguish users from each other. Therefore, approaches based on machine learning are required for user identification based on typing patterns.

We consider the task of person identification based on dynamics of typing as a time-series classification problem, for which various approaches have been introduced ranging from neural networks [6, 7] over Hidden Markov Models [8] to support vector machines [9] and Bayesian networks [10]. The 1-nearest neighbour (1NN) classifier with dynamic time warping (DTW) as distance measure was shown to be an extremely competitive classifier, outperforming many complex models, such as neural networks, Hidden Markov Models or "super-kernel fusion scheme" [11, 12]. While the empirical evidence is also justified by theoretical results [13, 14], one of the recently observed shortcomings of nearest neighbour models is their suboptimal performance in the presence of bad hubs [15, 16]. Informally, we say that an instance x is a *bad hub*, if x appears as the nearest neighbour of surprisingly many other instances, but x belongs to a class which is different from the class of those instances that have x as their nearest neighbour. With *hubness*, we refer to the presence of bad hubs, a phenomenon that has been observed in various datasets, including time series datasets [17]. For a more formal definition of bad hubs, we refer to [17], in which hubness-aware classifiers are surveyed and applied to the classification of time series. As the studies mentioned above show, bad hubs are responsible for a surprisingly large fraction of the total classification error of nearest neighbour classifiers, therefore, reduction of the detrimental effect of bad hubs can substantially improve the accuracy of time-series classification.

The dynamics of typing can be described by time series and hubness-aware models are among the most promising recent machine learning techniques for time-series classification. However, the presence of hubs has not been described in the context of data describing the dynamics of typing. Moreover, hubness-aware techniques have not been applied to person identification based on the dynamics of typing.

In this paper, we describe the presence of hubs in typing data and propose to use hubness-aware models for the task of person identification based on time series describing the dynamics of typing. We performed experiments on real-world data and show that hubness-aware models outperform prominent time-series classifiers. In order to be able to evaluate our approach, we collected data over several months from different users. In order to assist reproducibility and to motivate further research, we made our data publicly available at `http://biointelligence.hu/typing.html`.

The rest of the paper is organised as follows: Sec. 2 describes the hubness phenomenon in typing dynamics data and reviews the ECkNN hubness-aware machine learning technique. Sec. 3 gives the details of the proposed approach, while Sec. 4 presents our experimental results. Finally, we conclude in Sec. 5.

2 Hubs in Data Describing the Dynamics of Typing

In order to quantitatively study the presence of hubs in data describing the dynamics of typing, we use $N_k(x)$ to denote how many times an instance x appears as one of the k-nearest neighbours of other instances.[1] Denoting the number of instances by n, the total number of k-nearest neighbour occurrences is kn. Therefore, $E[N_k(x)] = k$. Note, however, that the nearest neighbour

[1]In our case, instances correspond to time series describing keystroke dynamics.

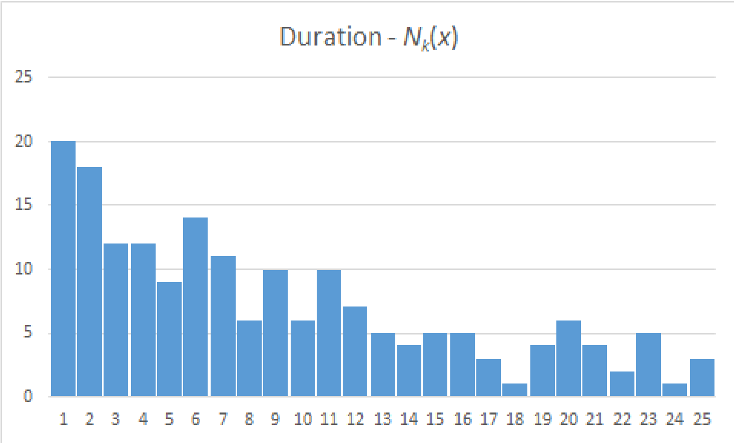

Figure 2 The distribution of $N_k(x)$ for the keystroke duration data. The horizontal axis shows the value of $N_k(x)$ with $k = 5$, while the vertical axis shows how many instances have that value of $N_k(x)$.

relationship is asymmetric (i.e., if an instance x is one of the nearest neighbours of x', this does not necessarily imply that x' is one of the nearest neighbours of x).

Therefore, as one can see in Fig. (2), $N_k(x)$ may largely vary for particular instances. The figure shows the distribution of $N_k(x)$ with $k = 5$ nearest neighbours for our keystroke duration data. We describe the data in more detail in Sec. 4. We used the aforementioned dynamic time warping distance to obtain the nearest neighbours. The horizontal axis of the figure shows the value of $N_k(x)$ while the vertical axis show how many instances have that value of $N_k(x)$. As one can see, there are instances that appear as nearest neighbours of surprisingly many other instances. For example, there are more than five instances that appear 20-times as nearest neighbours of other instances. Please note that the horizontal axis is truncated at $N_k(x) = 25$ for clarity, however, there are in total 17 instances that appear as nearest neighbours of *more* than 25 other instances. Similar observations can be made for the data containing the length of the time between consecutive keystrokes, see Fig. (3). The term *hubness* is used to refer to the presence of hubs.

Hubness, as it was described above, is not necessarily bad for classification, however, some of the hubs are bad in the sense that their class labels mismatch the class labels of those instances that have them as one of their nearest neighbours. More precisely, we may define the concept of *good* and *bad neighbors* as follows: an instance x is a good (bad, resp.) k-nearest neighbor of another instance x' if (i) x is one of the k-nearest neighbors of x' and (ii) they have the same (different, resp.) class labels. We use $GN_k(x)$ and $BN_k(x)$ to denote how many times an instance x appears as good and bad nearest neighbours of other instances. We say that an instance x is a bad hub if it $BN_k(x)$ value is exceptionally large. Both the distribution of $GN_k(x)$ and $BN_k(x)$ are similar to the distribution of $N_k(x)$, thus, there are bad hubs in the data.

Bad hubs were shown to be responsible for a surprisingly high fraction of the overall classification error [15]. Therefore, the presence of them should be taken into account. Hubness-aware techniques, such as the ECkNN, which is introduced in the next section, aim to reduce the

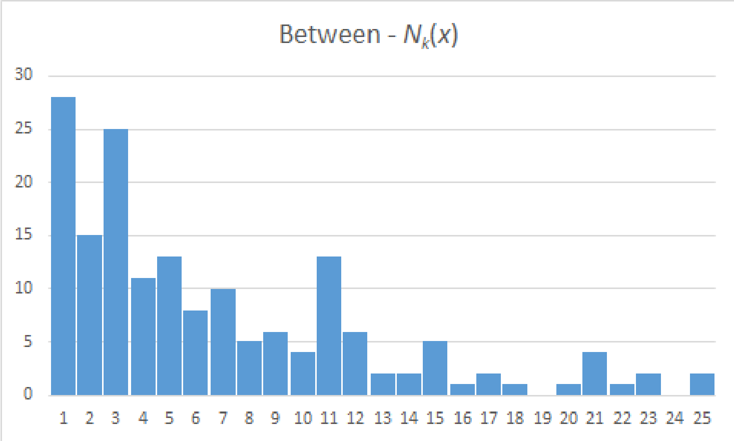

Figure 3 The distribution of $N_k(x)$ for the data containing the length of the time between consecutive keystrokes. The horizontal axis shows the value of $N_k(x)$ with $k = 5$, while the vertical axis shows how many instances have that value of $N_k(x)$.

detrimental effect of bad hubs.

2.1 Nearest neighbour regression with error correction

Nearest neighbour regression with error correction (ECkNN) is a hubness-aware extension of the kNN regression. By design, it is suitable for various types of data, e.g. vector data, time series, etc., given that an appropriate distance measure between the instances of the dataset is available. As we work with time-series data describing the dynamics of typing, the instances are time series in our case. Next, we describe ECkNN in more detail.

In its training phase, ECkNN implements error correction on the training data. In particular, the corrected label $y_c(x)$ of an instance x is defined as

$$y_c(x) = \begin{cases} \frac{1}{|\mathcal{I}_x|} \sum\limits_{x_i \in \mathcal{I}_x} y(x_i) & \text{if } |\mathcal{I}_x| \geq 1 \\ y(x), & \text{otherwise} \end{cases}, \tag{1}$$

where $\mathcal{I}_x$ denotes the set of training instances that have x as one of their k-nearest neighbours and $y(x)$ is the original (i.e., uncorrected) label of instance x. When ECkNN is applied to predict labels for new instances, it performs k-nearest neighbour regression using the corrected labels. That is: for a new instance $x*$, ECkNN searches for the k-nearest neighbours of $x*$ among the training instances and outputs the average of the corrected labels of the neighbours as the estimated label of $x*$.

Next, we illustrate the error correction mechanism performed by ECkNN on a simple example shown in Fig. (4). In the figure, training instances are denoted by circles. They are identified by the symbols $x_1 \dots x_7$. The numeric value (0 or 1) next to each instance shows its label. In order to keep the example simple, we use $k = 1$ nearest neighbour to calculate the corrected labels of training instances. In the figure, directed edges point from each instance to its first nearest

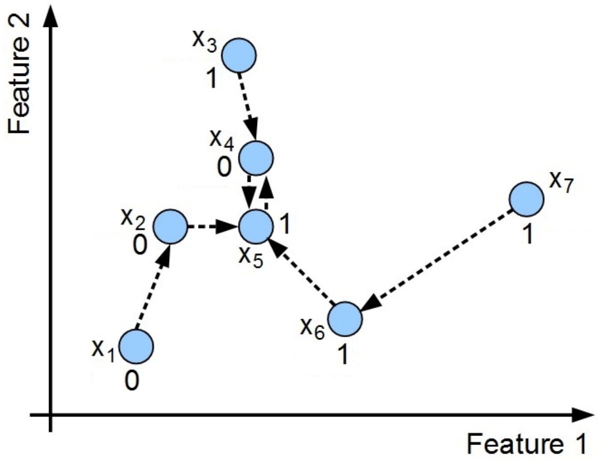

Figure 4 Example used to illustrate error correction.

neighbour. We only present the calculations for x_4 and x_5 as the procedure is similar in the case of the other instances as well. Concretely, the corrected labels of x_4 and x_5 are $y_c(x_4) = \frac{1}{2}(1+1) = 1$ and $y_c(x_5) = \frac{1}{3}(0+0+1) = 0.33$. For more details about ECkNN we refer to [18].

As mentioned previously, the dynamics of typing is captured by time series data. In order to use ECkNN with time series data, we need to be able to determine the nearest neighbours of time series. Dynamic time warping (DTW) is a time series distance measure that is robust to elongations and noise. DTW was originally introduced by Sakoe and Chiba [19]. In the last decades, DTW emerged as one of the most prominent techniques in machine learning with time series. Therefore, we use DTW as a distance measure for person identification based on time series describing the dynamics of typing. For a detailed description of DTW, we refer to [17].

3 Hubness-aware regression for person identification based on typing pattern

We base our solution on the wide-spread "classification-via-regression" approach. In particular, we use ECkNN regression for the person identification task in the following way: for each pair of users (u, v) we train a separate model. While doing that, we associate time series describing the typing dynamics of user u with label "0". Similarly, the time series of user v are associated with label "1". When a new time series is presented to the trained model, the model outputs a continuous value between 0 and 1 (bounds are inclusive). Values close to 0 (or 1, resp.) indicate that, according to the model, the new time series is more likely to represent the typing dynamics of user u (or v, respectively).

Usage of pairwise models, i.e., models that distinguish between two users, is consistent with the circumstances under which the person identification problem arise in real-world applications. Both in the case of the aforementioned online exams and online banking scenario, the user claims an identity and the task is to decide if the claimed identity matches the user's true identity. For example, if the claimed identity is u^*, and there are other users $u_1, u_2, \ldots, u_n$ in the system, we decide (i) if the typing pattern is more consistent with the typing patterns of u^* than u_1, and (ii) if the presented typing pattern is more consistent with the typing patterns of u^* than u_2, etc. These n decisions can be implemented in parallel if the number of users is high and several computational units are available.[2]

At each of the above decisions, a simple decision threshold of 0.5 could be applied. It means that given a model trained to distinguish between the typing patterns of users u and v, if the model outputs less than 0.5 when a new time series is presented to the model, then the decision is u; otherwise the decision is v. However, the simple threshold of 0.5 may be suboptimal. Therefore, we learn the threshold in the following way: once the model is trained, we present the time series of the training set to the model and obtain the output of the model for the training time series. Then, we determine the threshold that gives the highest accuracy on the training data.

4 Experimental evaluation

We begin this section by describing the data we used in our experiments. Subsequently, we provide the details of our experimental protocol and the results of the experiments we performed.

4.1 Typing Dynamics Data

We collected time series data describing the dynamics of typing, or *typing patterns* for short, from 12 different users over several months, resulting in a collection of 548 typing patterns in total. In each of the typing sessions, the users were asked to type the same short text of a few sentences. In particular, the users were asked to type the following text based on the English Wikipedia page about Neil Armstrong:

That's one small step for a man, one giant leap for mankind. Armstrong prepared his famous epigram on his own. In a post-flight press conference, he said that he decided on the words just prior to leaving the lunar module.

In each typing session, we measured both (i) the time between consecutive keystrokes and (ii) the duration of each keystroke, i.e., the time between pressing and releasing a key. We used a self-made JavaScript application and a PHP script to capture the aforementioned time series and to save the data. Note that due to typing errors, the length of typing patterns varies slightly from session to session. In order to encourage future research, we made our raw data publicly available and announced an open challenge at `http://biointelligence.hu/typing-challenge`.

4.2 Evaluation of pairwise models

In order to simulate the scenario in which users provide few typing patterns when they register into a system, we used the first five typing patterns per user as training data. The remaining typing patterns were used as test data in order to evaluate the system.

[2]As there might be users with similar typing dynamics, depending on the costs of different types of errors, in order to successfully authenticate user u^* we might allow a few of these pairwise decisions to "fail" in the sort of

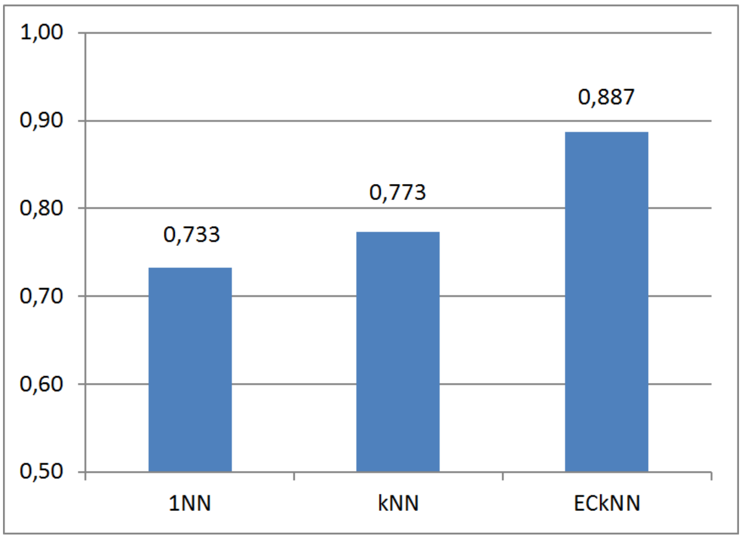

Figure 5 Accuracy of our approach (ECkNN) and the baselines.

For the reasons mentioned in Sec. 2, we trained models to distinguish two users, i.e., we trained a model for each pair of users.

For each model, we measured the accuracy, i.e., the ratio of correctly classified instances. We report accuracies averaged over all the pairs of users. We used the binomial test suggested by Salzberg [20] to judge if the differences between the models are statistically significant or not.

We measured the accuracy of our approach, ECkNN, and the baselines. We used both the time series of the times between consecutive keystrokes, and the time series of the durations of each keystroke. In the case of our approach, as well as in the case of the baselines, for both types of time series, we trained a separate model and combined the output of the two models by averaging their outputs.

We used the public ECkNN implementation from the PyHubs library [21]. We set $k = 5$ for ECkNN which is in accordance with other works on hubness-aware machine learning [18, 22].

As described in Section I, 1NN-DTW was reported as an extremely competitive time series classifier. Therefore we used it as one of the baselines. Additionally, we used k-nearest neighbour regression (kNN) with DTW and k=5.

Fig. (5) summarizes the results of the evaluation of pairwise models. Fig. (5) shows classification accuracy of the combined models that use both types of time series. We observed that ECkNN outperforms both baselines statistically significantly according to binomial tests at the significance level of $p = 0.001$.

Regarding the performance of both types of time series, we observed that keystroke duration time-series are more informative than the times between consecutive keystrokes.

Furthermore, we note that there is a non-negligible difference between the users regarding classification performance. On the one hand, most of the users can be distinguished well from each other, as they have unique typing dynamics. On the other hand, typing dynamics of some

sense that the model outputs that the typing pattern is more likely to belong to user u_i instead of u^*.

users varies remarkably. In the case of our data, the user numbered 4 was the most difficult to distinguish from other users. Such users may be difficult to recognise, and they may be able to imitate other users successfully. It is in line with the observations made by Doddington et. al [23]. In terms of Doddington's zoo, we observed the presence of "sheep", i.e., users on which the system works well, "goats" and "wolves", i.e., users that are difficult to recognise and users who may be successful in imitating others. While we believe that these observations are interesting, we note that a detailed analysis of keystroke dynamics data in terms of Doddington's zoo is left for future work.

4.3 Person Authentication Experiments

We performed additional experiments in the user authentication scenario using all the collected data. With *user authentication* we mean that the user claims an identity and the task is to decide whether the claimed identity matches the true identity. Similarly to the previous experiment, for each user, we used the first five typing patterns as training data, while the remaining typing patterns were used as test data. In order to simulate the scenario in which the user claims an identity, for the typing patterns in the test set, we used a random generator to obtain a *hypothetical identity* which matched the true user identity in 50 % of the cases.

Assume that the typing pattern p is associated with the hypothetical user identity u'. In this case, all the pairwise models deciding between the pairs of users (u_1, u'), (u_2, u'), ... made predictions for the user associated with the typing pattern p and they vote according to their predictions for or against the hypothetical identity. If there are more votes against the hypothetical identity than a certain threshold, we say the authentication failed, i.e., the hypothetic identity does not match the true identity. We call the aforementioned threshold a tolerance parameter and set it to two in our experiments.[3]

Our results show that our approach outperforms the baselines in the person authentication scenario as well, see Fig. (6).

5 Conclusions and outlook

In this paper, we considered the task of person identification based on the dynamics of typing. We proposed to use a hubness-aware regression technique, ECkNN, for the person identification task. We compared the results of ECkNN with 1NN-DTW and kNN regression which are highly competitive baselines. In order to assist reproducibility and to encourage future research, we made our data publicly available and announced an open challenge at `http://biointelligence.hu/typing-challenge`.

Additionally to ECkNN, we tried further hubness-aware regression techniques, namely, EWkNN, error-based weighting with kNN regression and EWCkNN which combines error-based weighting and error correction. As these approaches gave similar results to the results with ECkNN, for brevity, we omit to present them.

[3]This setting of the tolerance threshold is in accordance with the expected value of votes against the claimed identity under the assumption that the claimed identity is the true identity.

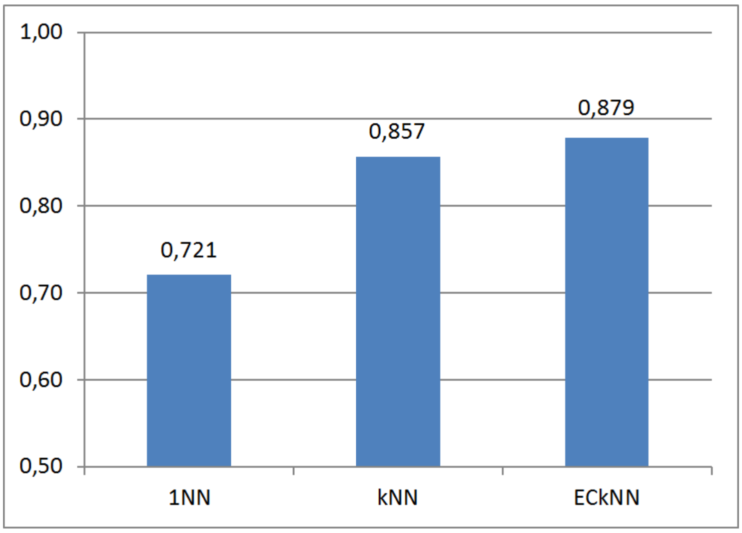

Figure 6 Accuracy of our approach and the baselines in the person authentication scenario.

With respect to the interpretation of the results, we note that we recorded typing patterns over a relatively long period of several months. As we used the first five typing patterns as training data and the remaining typing patterns as test data, the high accuracy we achieved indicates that, for the users we examined, the dynamics of typing was relatively stable over time.

In the underlying application, a relatively small set of reliably labelled data is given, e.g. the typing patterns that were recorded when the user registered to the system. At the same time, substantially more unlabeled data (or "weakly labelled" data, under the assumption that the user claims an identity when she tries to log in) may be collected during the usage of the system. The presence of unlabeled data was not taken into account in our work, but it might be exploited using semi-supervised machine learning techniques, such as the SUCCESS approach [24] that was designed for time series classification.

As hubness-aware approaches performed well for the person identification based on the dynamics of typing, we envision that similar techniques could be applied to person identification based on electroencephalograph (EEG) and electrocardiograph (ECG) signals as well.

Acknowledgemnts

We thank Ladislav Peška for his comments and remarks on the manuscript. This research was performed within the framework of the grant of the Hungarian Scientific Research Fund – OTKA PD 111710. This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Dóra Neubrandt was supported by the "Új Nemzeti Kiválóság Program" (ÚNKP-16-1-1) of the "Emberi Erőforrások Minisztériuma".

References

[1] S. Marcel and J.R. Del Millan. Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):743–752, 2007. DOI: 10.1109/TPAMI.2007.1012.

[2] F. Gargiulo, A. Fratini, M. Sansone, and C. Sansone. Subject identification via ecg fiducial-based systems: Influence of the type of QT interval correction. *Comput. Meth. Programs Biomed.*, 121(3):127–136, 2015. DOI: 10.1016/j.cmpb.2015.05.012.

[3] M. Fraschini, M. Demuru, L. Didaci, and G. Marcialis. An EEG-based biometric system using eigenvector centrality in resting state brain networks. *IEEE Signal Process. Lett.*, 22(6):1, 2015. DOI: 10.1109/LSP.2014.2367091.

[4] M. Antal, L.Z. Szabó, and I. László. Keystroke dynamics on Android platform. pages 820–826, 2015. DOI: 10.1016/j.protcy.2015.02.118.

[5] F. Monrose and A.D. Rubin. Keystroke dynamics as a biometric for authentication. *Futur. Gener. Comp. Syst.*, 16(4):351–359, 2000. DOI: 10.1016/S0167-739X(99)00059-X.

[6] F.W.M.H Wong, A.S.M. Supian, A.F. Ismail, L.W. Kin, and O.C. Soon. Enhanced user authentication through typing biometrics with artificial neural networks and k-nearest neighbor algorithm. In *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers, 2001.*, volume 2, pages 911–915. IEEE, 2001.

[7] A. Nanopoulos, R. Alcock, and Y. Manolopoulos. *Feature-based classification of time-series data*, pages 49–61. Nova Science Publishers, Commack, NY, USA, 2001.

[8] S. Kim, P. Smyth, and S. Luther. Modeling waveform shapes with random effects segmental hidden markov models. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 309–316. AUAI Press Arlington, Virginia, USA, 2004.

[9] D.R. Eads, D. Hill, S. Davis, S.J. Perkins, J. Ma, R.B. Porter, and J.P. Theiler. Genetic algorithms and support vector machines for time series classification. In *International Symposium on Optical Science and Technology*, pages 74–85. International Society for Optics and Photonics, 2002.

[10] K. Buza and L. Schmidt-Thieme. Motif-based classification of time series with Bayesian Networks and SVMs. In *Advances in Data Analysis, Data Handling and Business Intelligence*, pages 105–114. Springer, 2009. DOI: 10.1007/978-3-642-01044-6_9.

[11] X. Xi, E. Keogh, C. Shelton, L. Wei, and C.A. Ratanamahatana. Fast time series classification using numerosity reduction. In *Proceedings of the 23rd international conference on Machine learning - ICML '06*, pages 1033–1040. ACM, New York, NY, USA, 2006. DOI: 10.1145/1143844.1143974.

[12] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment*, 1(2):1542–1552, 2008. DOI: 10.14778/1454159.1454226.

[13] G.H. Chen, S. Nikolov, and D. Shah. A latent source model for nonparametric time series classification. In *Advances in Neural Information Processing Systems*, pages 1088–1096, 2013.

[14] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition*, volume 31 of *Stochastic Modelling and Applied Probability.* Springer Science & Business Media, 1996. DOI: 10.1007/978-1-4612-0711-5.

[15] M. Radovanović, A. Nanopoulos, and M. Ivanović. Hubs in space: Popular nearest neighbors in high-dimensional data. *J. Mach. Learn. Res.*, 11:2487–2531, 2010.

[16] N. Tomašev, M. Radovanovic, D. Mladenic, and M. Ivanovic. A probabilistic approach to nearest-neighbor classification: Naive hubness bayesian kNN. In *Proceedings of the 20th ACM international conference on Information and knowledge management - CIKM '11*, pages 2173–2176. ACM New York, NY, USA, 2011. DOI: 10.1145/2063576.2063919.

[17] N. Tomašev, K. Buza, K. Marussy, and P.B. Kis. Hubness-aware classification, instance selection and feature construction: Survey and extensions to time-series. In *Feature selection for data and pattern recognition*, volume 584 of *Studies in Computational Intelligence*, pages 231–262. Springer, Berlin-Heidelberg, Germany, 2015. DOI: 10.1007/978-3-662-45620-0_11.

[18] K. Buza, A. Nanopoulos, and G. Nagy. Nearest neighbor regression in the presence of bad hubs. *Knowledge-Based Syst.*, 86:250–260, 2015. DOI: 10.1016/j.knosys.2015.06.010.

[19] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust., Speech, Signal Process.*, 26(1):43–49, 1978. DOI: 10.1109/TASSP.1978.1163055.

[20] S.L. Salzberg. On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Min. Knowl. Discov.*, 1(3):317–328, 1997.

[21] PyHubs Software Package. `http://biointelligence.hu/pyhubs`, Accessed: 15.12.2016.

[22] N. Tomašev and D. Mladenić. Nearest neighbor voting in high-dimensional data: Learning from past occurrences. In *2011 IEEE 11th International Conference on Data Mining Workshops*, pages 1215–1218. IEEE, 2011. DOI: 10.1109/ICDMW.2011.127.

[23] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheep, goats, lambs and wolves a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. In *International Conference on Spoken Language Processing*, 1998.

[24] K. Marussy and K. Buza. SUCCESS: a new approach for semi-supervised classification of time-series. In *Artif. Intell. Soft. Comput.*, pages 437–447. Springer, 2013. DOI: 10.1007/978-3-642-38658-9_39.