Trusting Skype: Learning the Way People Chat for Fast User Recognition and Verification

Giorgio Roffo[1]
giorgio.roffo.it@ieee.org

Marco Cristani[1,2]
marco.cristani@univr.it

Loris Bazzani[2]
loris.bazzani@iit.it

Hà Quang Minh[2]
minh.haquang@iit.it

Vittorio Murino[1,2]
vittorio.murino@iit.it

[1] Department of Computer Science, University of Verona,
Strada le Grazie, 15, 37134, Verona

[2] Pattern Analysis & Computer Vision, Istituto Italiano di Tecnologia,
via Morego, 30, 16163 Genova

Abstract

Identity safekeeping on chats has recently become an important problem on social networks. One of the most important issues is identity theft, where impostors steal the identity of a person, substituting her in the chats, in order to have access to private information. In the literature, the problem has been addressed by designing sets of features which capture the way a person interacts through the chats. However, such approaches perform well only on the long term, after a long conversation has been performed; this is a problem, since in the early turns of a conversation, much important information can be stolen. This paper focuses on this issue, presenting a learning approach which boosts the performance of user recognition and verification, allowing to recognize a subject with considerable accuracy. The proposed method is based on a recent framework of one-shot multi-class multi-view learning, based on Reproducing Kernel Hilbert Spaces (RKHS) theory. Our technique reaches a recognition rate of 76.9% *in terms of AUC of the Cumulative Matching Characteristic curve, with only* 10 *conversational turns considered, on a total of* 78 *subjects. This sets the new best performances on a public conversation benchmark.*

1. Introduction

In recent years, cyber-attacks have become increasingly common , with many types of strategies and heterogeneous targets. In this panorama, one fact has emerged clearly: the interest of online criminals and spammers in using email as infection vectors has decreased dramatically [7]. Instead, their attention focuses now on social media, which represent a means with two appealing characteristics: social proof and sharing. Social proofing is the psychological mechanism that convinces people to do things because their friends are doing them [16]. Sharing is what people do with social networks: they share personal information such as their birthday, home address, and other critical information like credit card numbers etc.. This type of information is clearly precious for criminals who are now concentrating on these social applications, designing methods to steal virtual identities on instant messaging platforms, and grab private data from the victim and their contacts. Essentially, there are three ways with which an identity can be violated: by *identity theft* [15], where an impostor is able to access the personal account, mostly due to Trojan horse keystroke logging programs [9]; by *social engineering* (*i.e.*, tricking individuals into disclosing login details or changing user passwords) [3]; the last way consists in creating a *fake identity*, that is, an identity which describes an invented person, or emulates another person [12].

Since communication through social networks, such as Facebook, Twitter, and Skype is rapidly growing [22], identity violation is becoming a primary threat to people's cultural attitudes and behaviours in social networking. To give some numbers, the Federal Trade Commission reported that 9.9 million (22% more than 2007) Americans suffered from identity theft in 2008 [11]. The urgency of attacking the identity violation problem drove several institutions (banks, enforcement agencies and judicial authorities) to produce strategies and methods capable of discovering as soon as possible potential threats: they have been collected under the umbrella of the Identity Theft Red Flags Rule, issued in 2007. These strategies should be triggered by patterns, practices, or specific activities, known as "red flags", that could indicate identity theft [11].

In this paper, we follow this line, investigating possible technologies aimed at revealing the genuine identity of a person involved in instant messaging activities. In practice, we require that the user under analysis (from now on, the *probe* individual) engages a conversation for a very limited number of turns, with whatever interlocutor: after that, our cues can be extracted, providing statistical measures which can be matched with a *gallery* of signatures, looking for possible correspondences. Subsequently, the matches can be employed for performing user recognition or verification.

In the literature, very few approaches deal with this problem: in [8], a solution for the recognition problem is proposed, on a dataset of 77 individuals; the verification is then added in [18]. Both works consider chats as hybrid entities, that is crossbreeds of literary text and spoken conversations. Following this intuition, two pools of mixed features have been presented, taking inspiration from both the literature of Authorship Attribution, which recognizes the authors of pieces of text [1], and the one of non-verbal conversation analysis, where the way speakers chat (using emoticons, answering promptly after the other's turn) is modeled [21]. These "stylometric" features do a valid job in recognizing people, but high accuracies are obtained using a high number of turns (around 60), averaging on the distances between the different features as matching criterion. This is highly impractical, since in a real situation people need to be recognized after a few turns, while in this case the state of the art reports low performances.

Our approach deals with this problem, assuming that we want to recognize a person after no later than 5-10 conversation turns. We meet this goal by modifying a recent multi-class classification approach [14], allowing to exploit stylometric features in a much more powerful manner. Roughly speaking, each class corresponds to the identity of one individual; moreover, the approach allows the exploitation of multiple features, independently of their nature. In particular, we exploit the general framework of multi-view (or feature) learning with manifold regularization in vector-valued Reproducing Kernel Hilbert Spaces (RKHS). In this setting, each feature is associated with a component of a vector-valued function in an RKHS. Unlike multi-kernel learning [4], all components of a function are forced to map in the same fashion, i.e., to distinguish in a coherent way the different individuals. The desired final output is given by their combination, in a form to be made precise below, which is a fusion mechanism joining together the different features.

In the remainder of this paper we first briefly review related approaches in Section 2. We then introduce in Section 3 our method, discussing its implementation and sketching the proposed multi-view learning framework. Experiments are reported in Section 4, and, finally, Section 5 draws some conclusions and future perspectives.

2. Related work

Authorship Attribution (AA) aims at automatically recognizing the author of a given text sample, based on the analysis of *stylometric* cues. AA attempts date back to the 15th century[19]: since then, many stylometric cues have been designed, usually partitioned into five major groups: *lexical, syntactic, structural, content-specific and idiosyncratic* [1]. In the recent work of [18], *turn-taking* features have been crafted. Table 1 is a synopsis of the features applied so far in the literature.

Typically, state-of-the-art approaches extract stylometric features from data and use discriminative classifiers to identify the author (each author corresponds to a class). The application of AA to chat conversations is recent (see [20] for a survey), with [23, 2, 13, 17] the most cited works. In [23], a framework for authorship attribution of online messages is developed to address the identity-tracing problem. Stylometric features are fed into SVM and neural networks on 20 subjects, validating the recognition accuracy on 30 random messages. PCA-like projection is applied in [2] for Authorship identification and similarity detection on 100 potential authors of e-mails, instant messages, feedback comments and program code. A unified data mining approach is presented in [13] to address the challenges of authorship attribution in anonymous online textual communication (email, blog, IM) for the purpose of cybercrime investigation. In [17], 4 authors of IM conversations have been identified based on his or her sentence structure and use of special characters, emoticons, and abbreviations.

The main limitation of the works above is that they do not process chat exchanges as conversations, but as normal texts. In practice, the feature extraction process is always applied to the entire conversation and individual turns, which, while being the basic blocks of the conversation, are never used as analysis unit. In [8, 18], these limits are overcome by designing features which analyze each single turn as basic entity, considering aspects from both the AA literature and the non-verbal conversational analysis. With respect to the state of the art, our work combines good feature extraction with a powerful learning framework, which is adapted for recognition and verification purposes. In particular, we define the problem as a multi-shot person re-identification task [5, 10], where multiple instances of an individual are used to model his identity. Whereas in the re-identification literature the instances are images of the individual, here they are turns of chat conversations.

Group	Description	Examples	References
Lexical	*Word level*	Total number of words (=M), # short words/M, # chars in words/C, # different words, chars per word, freq. of stop words	[2, 13, 17, 20, 23]
	Character level	Total number of characters (chars) (=C), # uppercase chars/C, # lowercase chars/C, # digit chars/C, freq. of letters, freq. of special chars	[2, 17, 20, 23]
	Character—Digit n-grams	Count of letter—digit n-gram (a, at, ath, 1 , 12 , 123)	[2, 20, 23]
	Word-length distribution	Histograms, average word length	[2, 13, 17, 20, 23]
	Vocabulary richness	Hapax legomena, dislegomena	[2, 13, 20, 23]
	Length n-grams	Considers solely the length of the words; xo_LT is the length n-gram of order x.	[18]
Syntactic	Function words	Frequency of function words (of, for, to)	[2, 13, 17, 20, 23]
	Punctuation	Occurrence of punctuation marks (!, ?, :), multiple !—?	[2, 13, 17, 20, 23]
	Emoticons—Acronym	:-), L8R, Msg, :(, LOL; emoticons categories such as *Positive* that counts the occurrences of happiness, love, intimacy, etc. icons (20 emot. types in total) ; *Negative*: address fear, anger, etc. (19 emot. types in total); and *Other* or neutral emoticons portray actions, objects etc. (62 emot. types in total)	[17, 20, 18]
Structural	Message level	Has greetings, farewell, signature	[2, 13, 17, 20, 23]
Content-specific	Word n-grams	Bags of word, agreement (ok, yeah, wow), discourse markers—onomatopee (ohh), # stop words, # abbreviations, gender—age-based words, slang words	[2, 13, 17, 20, 23]
Idiosyncratic	Misspelled word	Belveier instead of believer	[2, 13, 17, 20]
Turn-taking	*Turn duration*	Time spent to complete a turn (in seconds);	[18]
	Writing speed	Number of typed characters or words per second;	[18]
	Answer Time	Time spent to answer a question in the previous turn of another interlocutor	[18]
	Mimicry	Ratio between number of chars -or words- in current turn and number of chars -or words- in previous turn of the opposite subject;	[18]

Table 1. Synopsis of the state-of-the-art features for AA on chats. "#" stands for "number of". In red we have the features that we used in our approach (best viewed in colors).

3. Our approach

The pipeline of the proposed approach is explained in the following. During the learning stage, training conversations of different subjects are collected to form the gallery set. The feature descriptors of each individual are extracted from the related conversations (*i.e.*, conversation in which he is involved), forming the user signature for that individual. Then, the similarity between the descriptors is computed for each feature by means of kernel matrices (see Sec. 3.1). Multi-view learning consists of estimating the parameters of the model given the training set (see Sec. 3.2). Given a probe signature, the testing phase consists of computing the similarity of each descriptor with the training samples and using the learned parameter to classify it (see Sec. 3.3).

3.1. Features and Kernels

In our work, we examine chats among pairs of people, *i.e.*, dialogic interactions. These conversations can be considered as sequences of *turns*, where each "*turn*" is a set of symbols typed consecutively by one subject without being interrupted by the other person. In addition, each turn is composed by one or more *sentences*: a sentence is a stream of symbols which is ended by a "return" character. Each sentence is labeled by a temporal ID reporting the time of delivery.

For each person involved in a conversation, we analyze his stream of turns (suppose T), ignoring the input of the other subject. This means that we assume that the chat style (as modeled by our features) is independent from the interlocutor - this assumption has been validated experimentally. From these data, a personal signature is extracted, that is composed of different cues, written in red in Table 1.

Our approach differs from the other standard AA approaches, where the features are counted over entire conversations, giving a single quantity. We consider the turn as a basic analysis unit, obtaining T numbers for each feature. For ethical and privacy issues, we discard any cue which involves the content of the conversation. Even if this choice is very constraining, because it prunes out many features of Table 1, the results obtained are very encouraging.

Given the descriptor, we extract a kernel from each feature. In particular, we used χ^2 kernels that they have been proved to perform well in practice in different applications.

3.2. Multi-view Learning

In this section, we briefly summarize the multi-view learning framework proposed in [14], with particular focus on user recognition in chats. We suppose to have for training a labeled gallery set $\{(x_i, y_i)\}$, where $x_i \in \mathcal{X}$ represents the i-th signature of the user with label (identity) $y_i \in \mathcal{Y}$.

Given that P is the number of identities in the re-identification problem, let the output space be $\mathcal{Y} = \mathbb{R}^P$. Each output vector $y_i \in \mathcal{Y}$, $1 \leq i \leq l$, has the form $y_i = (-1, \ldots, 1, \ldots, -1)$, with 1 at the p-th location if x_i is in the p-th class. Let m be the number of views/features and $\mathcal{W} = \mathcal{Y}^m = \mathbb{R}^{Pm}$.

We define user recognition as the following optimization problem based on the least square loss function:

$$f^\star = \operatorname{argmin}_{f \in \mathcal{H}_K} \frac{1}{l} \sum_{i=1}^{l} ||y_i - Cf(x_i)||^2_{\mathcal{Y}} + \gamma_A ||f||^2_{\mathcal{H}_K} + \gamma_I \langle \mathbf{f}, M\mathbf{f} \rangle_{\mathcal{W}^l}, \quad (1)$$

where

- f is a vector-valued function in an RKHS $\mathcal{H}_K$ that is induced by the matrix-valued kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^{Pm \times Pm}$, with $K(x, t)$ being a matrix of size $Pm \times Pm$ for each pair $(x, t) \in \mathcal{X} \times \mathcal{X}$,
- $f(x) = (f^1(x), \ldots, f^m(x))$, where $f^i(x) \in \mathbb{R}^P$ is the value corresponding to the ith view,
- $\mathbf{f} = (f(x_1), \ldots, f(x_l))$ as a column vector in $\mathcal{W}^l$,
- C is the combintation operator that fuses the different views as $Cf(x) = \frac{1}{m}(f^1(x) + \cdots + f^m(x)) \in \mathbb{R}^P$,
- $\gamma_A > 0$ and $\gamma_I \geq 0$ are the regularization parameters,
- M is defined as $M = I_l \otimes (M_m \otimes I_P)$, where $M_m = mI_m - \mathbf{e}_m \mathbf{e}_m^T$ [14].

The first term of Eq. 1 is the least square loss function that measures the error between the estimated output $Cf(x_i)$ for the input x_i with the given output y_i for each i. Given an instance x with m features, $f(x)$ represents the output values from all the features, constructed by their corresponding hypothesis spaces, that are combined through the combination operator C. The second term is the standard RKHS regularization term. The last term is the multi-view manifold regularization [14], that performs consistency regularization across different features.

The solution of the minimization problem of Eq. 1 is unique [14]: $f^\star(x) = \sum_{i=1}^{l} K(x, x_i)\, a_i$, where the vectors a_i are given by the following system of equations:

$$(\mathbf{C}^*\mathbf{C}K[\mathbf{x}] + l\gamma_I MK[\mathbf{x}] + l\gamma_A I)\mathbf{a} = \mathbf{C}^*\mathbf{y}, \quad (2)$$

where $\mathbf{a} = (a_1, \ldots, a_l)$ is a column vector in $\mathcal{W}^l$ and $\mathbf{y} = (y_1, \ldots, y_l)$ is a column vector in $\mathcal{Y}^l$. Here $K[\mathbf{x}]$ denotes the $l \times l$ block matrix whose (i, j) block is $K(x_i, x_j)$;

$\mathbf{C}^*\mathbf{C}$ is the $l \times l$ block diagonal matrix, with each diagonal block being C^*C; $\mathbf{C}^*$ is the $l \times l$ block diagonal matrix, with each diagonal block being C^*.

Assume that each input x is decomposed into its m different views, $x = (x^1, \ldots, x^m)$. For our setting, the matrix-valued kernel $K(x, t)$ is defined as a block diagonal matrix, with the (i, i)-th block given by

$$K(x, t)_{i,i} = k^i(x^i, t^i) I_P, \quad (3)$$

where k^i is a kernel of the i-th views as defined in Sec. 3.1. To simply the computation of the solution we define the matrix-valued kernel $G(x, t)$, which for each pair $(x, t) \in \mathcal{X} \times \mathcal{X}$ is a diagonal $m \times m$ matrix, with

$$(G(x, t))_{i,i} = k^i(x^i, t^i), \quad (4)$$

The Gram matrix $G[\mathbf{x}]$ is the $l \times l$ block matrix, where each block (i, j) is the respective $m \times m$ matrix $G(x_i, x_j)$. The matrix $G[\mathbf{x}]$ then contains the Gram matrices $k^i[\mathbf{x}]$ for all the kernels corresponding to all the views. The two matrices $K[\mathbf{x}]$ and $G[\mathbf{x}]$ are related by

$$K[\mathbf{x}] = G[\mathbf{x}] \otimes I_P. \quad (5)$$

The system of linear equations 2 is then equivalent to

$$BA = Y_C, \quad (6)$$

where

$$B = \left(\frac{1}{m^2}(I_l \otimes \mathbf{e}_m \mathbf{e}_m^T) + l\gamma_I (I_l \otimes M_m) \right) G[\mathbf{x}] + l\gamma_A I_{lm},$$

which is of size $lm \times lm$, A is the matrix of size $lm \times P$ such that $\mathbf{a} = \operatorname{vec}(A^T)$, and Y_C is the matrix of size $lm \times P$ such that $\mathbf{C}^*\mathbf{y} = \operatorname{vec}(Y_C^T)$.

Solving the system of linear equations 6 with respect to A is equivalent to solving system 2 with respect to $\mathbf{a}$.

3.3. Testing

The testing phase consists of computing $f^\star(v_i) = \sum_{j=1}^{l} K(v_i, x_j) a_j$, given the testing set $\mathbf{v} = \{v_1, \ldots, v_t\} \in \mathcal{X}$. Let $K[\mathbf{v}, \mathbf{x}]$ denote the $t \times l$ block matrix, where block (i, j) is $K(v_i, x_j)$ and similarly, let $G[\mathbf{v}, \mathbf{x}]$ denote the $t \times l$ block matrix, where block (i, j) is the $m \times m$ matrix $G(v_i, x_j)$. Then

$$\mathbf{f}^\star(\mathbf{v}) = K[\mathbf{v}, \mathbf{x}]\mathbf{a} = \operatorname{vec}(A^T G[\mathbf{v}, \mathbf{x}]^T).$$

For the i-th sample of the p-th user, $f^\star(v_i)$ represents the vector that is as close as possible to $y_i = (-1, \ldots, 1, \ldots, -1)$, with 1 at the p-th location. The identity of the i-th image is estimated *a-posteriori* by taking the index of the maximum value in the vector $f^\star(v_i)$.

4. Experiments

In the experiments, we consider a dataset of Skype conversations, available at `http://profs.sci.univr.it/~cristanm/code.html`, which is explained in the following. First of all, we performed identity recognition in order to investigate the ability of the system in recognizing a particular probe user among the gallery subjects. To this end, we consider conversations which are 10 turns long, i.e., very short dyads, in order to modulate the number of training conversations that we can have for each individual. Then, keeping fixed the number of training conversations for each user (3 conversations), we vary the number of turns from 2 to 10 to test the accuracy of the proposed method using a limited number of turns. After this, we analyze the user verification: the verification performance is defined as the ability of the system in verifying if the person that the probe user claims to be is truly him/herself, or if he/she is an impostor.

As a performance measure for the identity recognition, we used the Cumulative Matching Characteristic (CMC) curve. The CMC is an effective performance measure for AA approaches [6]: given a test sample, we want to discover its identity among a set of N subjects. In particular, the value of the CMC curve at position 1 is the probability (also called *rank1* probability), that the probe ID signature of a subject is closer to the gallery ID signature of the same subject than to any other gallery ID signature; the value of the CMC curve at position n is the probability of finding the correct match in the first n ranked positions. As a single measure to summarize a CMC curve, we use the normalized Area Under the Curve (nAUC), which is the approximation of the integral of the CMC curve. For the identity verification task, we report the standard ROC curves, together with the Equal Error Rate (EER) values.

As a comparative approach, we consider the strategy of [18], whose code is available at the same page of the dataset.

4.1. The dataset

The corpus of [18] consists of 312 dyadic Italian chat conversations collected with Skype, performed by $N = 78$ different users [1]. The conversations are spontaneous, i.e. they were held by the subjects in their real life, collected over a time span of 5 months: in particular, for each individual there are around 13 hours of chatting activity. The number of turns per subject ranges between 200 and 1000. Our experiments are performed over at most 4 conversations of each person, in order to have the same number of conversations for all the people in the dataset. The conversations of each subject are split into *probe* and *gallery* sets, where the *probe* include just one conversation made of $TT = 10$ turns, the *gallery* can include from 1 to 3 conversations where each of them is again made of 10 turns. In this way, any bias due to differences in the amount of available material is avoided. When possible, we pick different conversations selections in order to generate different probe/gallery partitions.

[1] Conversation are intended in [18] as consecutive exchanges of turns with an interval between them not superior to 30 minutes.

ID	Name	Range
1	#Words(W)	[0,1706]
2	#Chars(C)	[0,15920]
3	Mean Word Length	[0,11968]
4	#Uppercase letters	[0,11968]
5	*#Uppercase / C*	[0,1]
6	*1o_LT*	[0,127]
7	*2o_LT*	[0,127]
8	# ? and ! marks	[0,21]
9	#Three points (...)	[0,54]
10	#Marks (",.:*;)	[0,1377]
11	*#Emoticons / W*	[0,4]
12	*#Emoticons / C*	[0,1]
13	*Turn Duration*	[0,1800]
14	*Word Writing Speed*	[0,562]
15	*Char Writing Speed*	[0,5214]
16	*#Emo. Pos.*	[0,48]
17	*#Emo. Neg.*	[0,5]
18	*#Emo. Oth.*	[0,20]
19	*Imitation Rate / C*	[0,2611]
20	*Imitation Rate / W*	[0,1128]
21	*Answer Time*	[0,2393]

Table 2. Stylometric features used in this work and related ranges of values assumed in our experiments.

In the Table 2 we report the features we used in our experiments, together with their ranges calculated on the entire dataset. For their meaning, we invite the reader to check Table 1, looking at the features colored in red.

4.2. Identity recognition

In the identity recognition task, we performed two experiments. In the first experiment we fixed the number of turns after which we want an answer from the system to $TT = 6$ (in the next experiment we varied this parameter also). After that, we built a training set, which for each person has a particular number of conversations, that is used by the learning algorithm to train the system. After training, we applied our approach on the testing set, which was composed of a conversation for each subject, performed the recognition, then calculated the CMC curve and the related nAUC value. We did the same with the comparative approach (which simply calculates distances among features, and computes the average distance among the probe conversation and the three training conversations). All the experiments were repeated 10 times, by shuffling the train-

Gallery Size	Roffo et al.[18] (nAUC)	Our approach (nAUC)
1 conv.	65.3%(8.9%)	**68.7%(10.0%)**
2 conv.	64.6%(10.7%)	**71.2%(11.4%)**
3 conv.	64.3%(11.1%)	**75.4%(12.6%)**

Table 3. Comparison between Roffo et al. [18] and the proposed method increasing the number of conversations (*conv.* formed by 6 turns each). The first number represents the nAUC, while in parenthesis we have the rank1 probability.

Turns	Roffo et al.[18]	Our approach
2	53.3%	**65.8%**
4	58.5%	**70.9%**
6	64.3%	**75.4%**
8	70.4%	**76.9%**
10	77.5%	**79.2%**

Table 4. Comparison between Roffo et al. [18] and the proposed method in terms of nAUC. We kept the number of conversation per subject in the gallery fixed, while we varied the number of turns per conversation.

ing/testing partitions. The results are better with our proposal both in case on nAUC and rank 1 score. In all the cases it is evident that augmenting the number of conversation gives a higher recognition score.

In the second experiment, we kept fixed the number of conversations per gallery to 3, and we gradually increased the number of turns from 2 to 10 with stepsize 2. The recognition results of [18] along with our method are reported in Table 4. It is easy to notice that our approach outperforms [18] in all the cases. and that the increment with respect to [18] is more evident when using a low number of turns. This result supports the fact that [18] is good only when having a lot of data and a good descriptor. Instead, the proposed approach can be used also with very few information.

4.3. Identity verification

Considering the verification task, we adopted the following strategy: given the signature of user i, if it matches with the right gallery signature with a matching distance which is ranked below the rank K, it is verified. Intuitively, there is a tradeoff in choosing K. A high K (for example, 78, all the subjects of the dataset) gives a 100% of true positive rate (this is obvious by construction), but it brings in a lot of potential false positives. Therefore, taking into account the number K as varying threshold, we can build ROC and precision/recall curves, using the value K as varying parameter to build the curves.

In particular, we report for each method, and for each number of turns taken into account: the nAUC of the ROC curve; the equal error rate (EER), which is the error rate occurring when the decision threshold of a system (K) is set so that the proportion of false rejections will be approximately equal to the proportion of false acceptances (less is better); the best F1 value obtained, together with the K value which gave the best results (in terms of F1 score), and the related precision and recall values. For the sake of the clarity, we produce two tables, one for the [18] method (Table 5), one for our approach (Table 6).

Turns	ROC (nAUC)	EER	Best F1 / Best K	Best Prec / Recall
2	52.8%	48.7%	67.3% / 4	51.7% / 96.2%
4	57.9%	45.4%	69.4% / 5	54.6% / 95.0%
6	63.8%	40.3%	70.5% / 5	55.9% / 95.1%
8	70.0%	34.2%	71.9% / 5	57.7% / 95.1%
10	77.3%	30.1%	74.7% / 10	64.3% / 88.9%

Table 5. Verification performance of Roffo et al. [18] approach.

Turns	ROC (nAUC)	EER	Best F1 / Best K	Best Prec / Recall
2	65.3%	39.4%	69.2% / 6	54.8% / 93.7%
4	70.5%	35.5%	70.2% / 6	56.1% / 93.8%
6	75.0%	32.6%	72.6% / 6	63.4% / 85.0%
8	76.6%	30.3%	73.2% / 9	61.7% / 90.1%
10	78.9%	28.9%	73.9% / 15	67.0% / 82.5%

Table 6. Verification performance of the proposed approach.

Our approach performs better, except in the case of 10 turns, where the F1 score is higher for [18]. It is worth noting that the higher F1 score is due to a very high recall score, definitely superior to the precision value. In our case, precision and recall are better balanced. In general, it is possible to note that the recall values are higher than the precision, and that augmenting the number of turns gives higher performances.

5. Conclusions

The ability to understand the identity of a person by looking at the way she chats is something that we can intuitively feel: we certainly know that some people are used to answering more promptly to our questions, or we know some people who are very fast in writing sentences. Our approach subsumes these abilities, putting them into a learning approach, which is capable of understanding the peculiar characteristics of a person, enabling effective recognition and verification. In particular, this study offers a first analysis of what a learning approach can do, when it comes to minimizing the information necessary to individualize a particular identity. The results are surprisingly promising: with just 2 turns of conversation, we are able to recognize and verify a person strongly above chance. This demonstrates that a form of behavioral blueprint of a person can be extracted even on a very small portion of chats. We believe therefore that our work has the potential to open up the possibility of a large range of new applications beyond surveillance and monitoring.

References

[1] A. Abbasi and H. Chen. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM TOIS*, 26(2):1–29, 2008.

[2] A. Abbasi, H. Chen, and J. Nunamaker. Stylometric identification in electronic markets: Scalability and robustness. *JMIS*, 25(1):49–78, 2008.

[3] R. J. Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems*. John Wiley & Sons, Inc., 2001.

[4] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan. Multiple kernel learning, conic duality, and the SMO algorithm. In *ICML*, 2004.

[5] L. Bazzani, M. Cristani, A. Perina, and V. Murino. Multiple-shot person re-identification by chromatic and epitomic analyses. *Pattern Recognition Letters*, 33(7):898–903, May 2012. Special Issue on Awards from ICPR 2010.

[6] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior. *Guide to Biometrics*. Springer Verlag, 2003.

[7] S. Corporation. *Internet Security Threat Report 2013*, volume 18. Symantec Publishing, April 2013.

[8] M. Cristani, G. Roffo, C. Segalin, L. Bazzani, A. Vinciarelli, and V. Murino. Conversationally-inspired stylometric features for authorship attribution in instant messaging. In *Proceedings of the 20th ACM international conference on Multimedia*, MM '12, pages 1121–1124, New York, NY, USA, 2012. ACM.

[9] Z. Deng, D. Xu, X. Zhang, and X. Jiang. Introlib: Efficient and transparent library call introspection for malware forensics. In *DFRWS*, pages 13 – 23, 2012.

[10] D. Figueira, L. Bazzani, M. H. Quang, M. Cristani, A. Bernardino, and V. Murino. Semi-supervised multi-feature learning for person re-identification. In *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 2013.

[11] K. M. Finklea. *Identity Theft: Trends and Issues*. DIANE Publishing, 2010.

[12] J. P. Harman, C. E. Hansen, M. E. Cochran, and C. R. Lindsey. Liar, liar: Internet faking but not frequency of use affects social skills, self-esteem, social anxiety, and aggression. *CB*, 8(1):1–6, 2005.

[13] F. Iqbal, H. Binsalleeh, B. C. M. Fung, and M. Debbabi. A unified data mining solution for authorship analysis in anonymous textual communications. *Information Sciences*, 2011.

[14] H. Q. Minh, L. Bazzani, and V. Murino. A unifying framework for vector-valued manifold regularization and multi-view learning. In *ICML*, volume 28, pages 100–108, 2013.

[15] R. C. Newman. Cybercrime, identity theft, and fraud: practicing safe internet - network security threats and vulnerabilities. In *InfoSecCD*, pages 68–78, 2006.

[16] Nielsen. *State Of Social Media: The Social Media Report 2012*. The Nielsen Company, April 12, 2012.

[17] A. Orebaugh and J. Allnutt. Classification of instant messaging communications for forensics analysis. *Social Networks*, pages 22–28, 2009.

[18] G. Roffo, C. Segalin, A. Vinciarelli, V. Murino, and M. Cristani. Reading between the turns: Statistical modeling for identity recognition and verification in chats. In *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2013)*, 2013.

[19] G. Shalhoub, R. Simon, R. Iyer, J. Tailor, and S. Westcott. Stylometry system–use cases and feasibility study. *Forensic Linguistics*, 1:8, 2010.

[20] E. Stamatatos. A survey of modern authorship attribution methods. *JASIST*, 60(3):538–556, 2009.

[21] A. Vinciarelli, M. Pantic, and H. Bourlard. Social Signal Processing: Survey of an emerging domain. *Image and Vision Computing Journal*, 27(12):1743–1759, 2009.

[22] D. L. Williams, V. L. Crittenden, T. Keo, and P. McCarty. The use of social media: an exploratory study of usage among digital natives. *JPA*, 2012.

[23] R. Zheng, J. Li, H. Chen, and Z. Huang. A framework for authorship identification of online messages: Writing-style features and classification techniques. *JASIST*, 57(3):378–393, 2006.