Predicting Authorship and Author Traits from Keystroke Dynamics

Barbara Plank
IT University of Copenhagen
Department of Computer Science
Rued Langgaards Vej 7, 2300 Copenhagen S, Denmark
bapl@itu.dk

Abstract

Written text transmits a good deal of non-verbal information related to the author’s identity and social factors, such as age, gender and personality. However, it is less known to what extent behavioral biometric traces transmit such information. We use typist data to study the predictiveness of authorship, and present first experiments on predicting both age and gender from keystroke dynamics. Our results show that the model based on keystroke features leads to significantly higher accuracies for authorship than the text-based system, while being two orders of magnitude smaller. For user attribute prediction, the best approach is to combine the two, suggesting that extra-linguistic factors are disclosed to a larger degree in written text, while author identity is better transmitted in typing behavior.

1 Introduction

Language is a social phenomenon (Nguyen et al., 2015). Whenever we speak or write we transmit a good deal of additional non-verbal information that is related to identity and social factors of an author. Early work in authorship analysis has typically been concerned with finding the author of a text, i.e., authorship attribution (Mosteller and Wallace, 1964; Stamatatos, 2009). In recent years, there has been a surge of interest towards the social dimension of language. Studies are interested in linking social factors with linguistic features, e.g., (Eisenstein et al., 2011; Bamman et al., 2014), studying data biases (Hovy and Søgaard, 2015) or building actual attribute prediction models from linguistic features (i.e., author profiling). Modeling author traits can further help to improve prediction of related attributes (Liu et al., 2016; Benton et al., 2017), help debiasing models (Hovy, 2015; Zhang et al., 2018) or can be used for a wide range of applications like customer support, healthcare and personalized machine translation (Mirkin et al., 2015; Rabinovich et al., 2017). Factors studied so far include gender, age, personality or income, to name but a few (Mairesse and Walker, 2006; Luyckx and Daelemans, 2008; Rao et al., 2010; Rosenthal and McKeown, 2011; Nguyen et al., 2011; Volkova et al., 2013; Flekova et al., 2016b; Verhoeven et al., 2016; van Dalen et al., 2017; Ljubešić et al., 2017; Emmery et al., 2017; van der Goot et al., 2018).

A key question in authorship analysis and profiling is what sorts of evidence might bear on determining authorship (Nerbonne, 2007) (or traits). What all prior work has in common is that it almost exclusively focused on the written text itself. As people read or write texts, they unconsciously produce cognitive by-product, such as gaze patterns or typist behavior. This evokes and motivates our research question: to what extent is behavioral data beyond the text predictive of authorship and author traits? In this paper we focus on keystroke dynamics. They concern a user’s typing pattern. Keystroke logs have the distinct advantage over other cognitive modalities like brain scans or gaze, that keystroke logs are more readily available; they do not rely on special equipment beyond a keyboard. While keystrokes are known to be informative for author verification (cf. Section 5), it is less clear to what extent keystrokes are predictive of authorship, and even more so, of author traits.

Contributions
a) We study the effect of keystrokes to identify authorship in two corpora of varying size. b) We investigate the predictive power of typist data for age and gender prediction. c) We compare behavioral measures to traditional stylometric features.
2 Keystroke dynamics

Keystroke logs are recordings of a user’s typing dynamics. When a person types on a keyboard, the latencies between successive keystrokes and their duration reflect the typing behavior of a person. For example, Figure 2 shows the keystroke hold times (average over single letters) of two users from our dataset. In its raw form, keystroke logs contain information on which key was pressed for how long (key, time press, time release). Research on keystroke dynamics typically consider timing measures derived from time press and time release events between keystrokes, such as key hold times or interkey durations (see Figure 1).

Figure 1: Keystroke logs illustrated: $p$ are pauses between keystrokes. Figure adapted from (Goodkind and Rosenberg, 2015).

Figure 2: Keystroke basic feature set: normalized mean key hold time for two example users.

Datasets The two keystroke datasets differ in the amount of users and available meta-data. The first, STEWART, stems from students taking a test on spreadsheet modeling (Stewart et al., 2011). This dataset is not distributed with further meta-data, hence it is used for authorship attribution only. The second dataset, VILLANI (Tappert et al., 2009), is larger (144 participants) and contains demographic meta-data. Keystrokes were recorded for two tasks: free text production and a copy task (fixed text snippet). As we are interested in author attribution/profiling, we consider only the former.

Pre-processing and Features First, we remove users with fewer than 5 typing sessions, sessions shorter than 5 words, users without demographics and users that only participated in the copy task (for VILLANI). We also removed two spammers (random skribble). This resulted in a dataset with 34 and 121 users with an average of 99 and 125 tokens per session for STEWART and VILLANI, respectively. The final gender/age distribution is not balanced: 53 female/68 male users, and 56 users above/65 user below thirty. For all keystrokes, the type of key was derived: letters, numbers, punctuation etc., ignoring control keys (FN etc).

Second, we derive 218 biometric features following (Stewart et al., 2011; Tappert et al., 2010). These biometric features include duration features (mean and standard deviation) and are grouped into: i) basic keystroke features, i.e., key hold time (key press and release time) features of the 26 letters from the English alphabet (cf. Figure 2 for an illustration); and ii) extended features: key hold times over groups of keys (like digits, punctuation etc) and transition (inter-key duration) features between successive keystrokes, e.g., between letters and non-letters, or individual letters and

3 Experiments

Given a dataset with keystroke logs, we run two sets of experiments: a) authorship attribution, i.e., to determine who wrote a given piece of text; and b) authorship profiling, i.e., to determine extralinguistic user traits, in particular age and gender.

Only very recently this source has been explored as information in natural language processing, for example, to aid shallow syntactic parsing (Plank, 2016) or deception detection (Banerjee et al., 2014) (Section 5). Keystroke logs have been used in computer security for user verification, however, combining keystroke biometrics with traditional stylometry metrics has not been proven successful (Stewart et al., 2011). The authors focused on a single task and dataset only. In contrast, in this paper we examine to what extent keystroke dynamics are informative for authorship attribution and author profiling.
Table 1: Results for authorship attribution (34 users), comparison of features (text vs keystrokes), and combined models. Best result in boldface.

<table>
<thead>
<tr>
<th></th>
<th>#Features</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (random)</td>
<td>–</td>
<td>0.4</td>
</tr>
<tr>
<td>Text</td>
<td>54k</td>
<td>50.2</td>
</tr>
<tr>
<td>Keystrokes (basic)</td>
<td>52</td>
<td>81.0</td>
</tr>
<tr>
<td>Keystrokes (extended)</td>
<td>218</td>
<td>90.2</td>
</tr>
<tr>
<td>Keystrokes (ext.) + Text</td>
<td>55k</td>
<td>70.2</td>
</tr>
<tr>
<td>Keystrokes (ext.) + Embeds</td>
<td>413</td>
<td>91.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VILLANI (121 users)</th>
<th>#Features</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>46k</td>
<td>67.1</td>
</tr>
<tr>
<td>Keystrokes (basic)</td>
<td>52</td>
<td>2.5</td>
</tr>
<tr>
<td>Keystrokes (ext.)</td>
<td>218</td>
<td>85.9</td>
</tr>
<tr>
<td>Keystrokes (ext.) + Text</td>
<td>46k</td>
<td>71.1</td>
</tr>
<tr>
<td>Keystrokes (ext.) + Embeds</td>
<td>413</td>
<td>88.1</td>
</tr>
</tbody>
</table>

4 Results

The results of training a classifier to predict the identity of an author are given in Table 1 and Figure 3. The random baseline accuracy is low (0.4% F1). Biometric behavioral features work incredibly well, reaching a performance in the 80-90ies. Already the basic feature set of 52 letter duration features clearly outperforms the stylistic features, reaching 81% F1-score. In contrast, stylistic features from the text alone reach an F1 of only 50%. Note that for the dataset with more users (VILLANI, Figure 3), results for authorship are actually higher, which may be explained by the fact that the smaller dataset is more controlled by topic (exam questions). Figure 3 shows that also on the larger dataset keystroke features outperform the text-based features (word and character n-grams) for authorship, even in setups with few users. These are remarkable results. The behavioral models employ a considerably smaller feature space (cf. column 2 in Table 1). Adding stylistic features improves performance over keystrokes, but only for the embeddings setup, which results in the best setup.

The results for author profiling are given in Table 2. Baseline results (majority baseline) are higher; this task is easier. The gap between stylistic and behavioral features is smaller, but the same trend holds: biometric behavioral features are predictive of gender. To a certain extent this also holds for age (albeit to a lesser extent). Interestingly, combining biometrics with traditional token-based features consistently proves the most effective for author profiling, albeit the best way differs per trait.

Our results suggest that author identity is highly captured in keystrokes alone, while the textual signal provides complementary evidence that together proves the most effective for predicting age and gender of an author.
5 Related Work

Authorship attribution has a long tradition dating back to early works in the 19th century. The most influential work on authorship attribution goes back to Mosteller and Wallace (1964). For a long time approaches to authorship attribution focused on distributions of function words, high-frequency words that are presumably not consciously manipulated by the author (Nerbonne, 2007; Pennebaker, 2011). Recent work also includes authorship studies on microblog texts (Schwartz et al., 2013). An recent survey is Stamatatos (2009). We here study another source of information that is presumably not consciously manipulated, keystroke dynamics.

A major scientific interest in keystroke dynamics arose in writing research, where it has developed into a promising non-intrusive method for studying cognitive processes involved in writing (Sullivan et al., 2006; Nottbusch et al., 2007; Wengelin, 2006; Van Waes et al., 2009; Baaijen et al., 2012). In these studies time measurements—pauses, bursts and revisions—are considered traces of the recursive nature of the writing process. Bursts are defined as consecutive chunks of text produced and defined by a 2000ms time of inactivity (Wengelin, 2006). In fact, most prior work that uses keystroke logs focuses on experimental research. For example, Hanoulle et al. (2015) study whether a bilingual glossary reduces the working time of professional translators. They consider pause durations before terms extracted from keystroke logs and find that a bilingual glossary reduces the translators’ workload. An analysis of users’ typing behavior was studied by Baba and Suzuki (2012) to measure the impact of spelling mistakes. Goodkind and Rosenberg (2015) investigate pre-word pauses and their relation to multi-word expressions. They found that within MWE pauses vary depending on the cognitive task. Banerjee et al. (2014) were the first to use keystroke patterns for deception detection.

Keystrokes were successfully used for author verification in computer security research (Stewart et al., 2011; Monaco et al., 2013; Locklear et al., 2014), as they are known to be idiosyncratic (Leggett and Williams, 1988). Our results show that keystroke biometrics are far superior over stylometry-based features in authorship attribution, and are predictive of author traits.

The study most related to ours (Stewart et al., 2011) used features from both keystrokes and linguistic stylometry for user verification in a k-nearest neighbor setup. Their study differs from ours in three aspects. First, they use a more elaborate set of stylometric features (like number of words of a certain length, and readability measures). Second, they target user authentication, thus their setup is a binary classification task (authenticated vs not-authenticated), while we here focus on a multi-class classification setup, which is a considerably more difficult task. Third, they use only a single dataset (STEWART), while we here include results on a second and larger dataset (n=121 authors). To the best of our knowledge, prior work on predicting demographics from typing behavior is typically limited to a single variable (Tsimperidis et al., 2015), except (Brizan et al., 2015), whose data is not available. Our study differs from theirs by studying age, and the focus on complementing textual with behavioral data.

Disclaimer While modeling user demographics can be seen as one step towards addressing biases in NLP it is important to be aware of potential negative side effects, both from the modeling side through potential exclusion or dual use (Hovy and Spruit, 2016), as well as the data side, when dealing with privacy sensitive data (cognitive behavioral data) or labels (e.g., mental health).

6 Conclusions

We have shown that behavioral biometrics contain highly predictive information for both authorship and author profiling. For authorship attribution, behavioral keystroke metrics significantly outperform traditional text-based features (words and character unigrams), while using a feature set which is orders of magnitude smaller (218 vs sev-
eral thousands of features). In addition, we show that keystroke dynamics are also predictive for author traits (gender and age). Interestingly, for the latter task, it is most beneficial to combine behavioral keystroke data with traditional text-based features, suggesting that user traits are disclosed to a larger degree in written text while identity is better disclosed in typing behavior.

Acknowledgments

I would like to thank the anonymous reviewers for their valuable feedback. This research is supported by NVIDIA corporation.

References


Kim Luyckx and Walter Daelemans. 2008. **Personae: a corpus for author and personality prediction from text.** In *LREC 2008*.


Frederick Mosteller and David Wallace. 1964. **Inference and disputed authorship: The Federalist.** Addison-Wesley.


Barbara Plank and Dirk Hovy. 2015. **Personality traits on twitter—or—how to get 1,500 personality tests in a week.** In *Proceedings of the 6th Workshop on*


