PAN 2017 — Digital Text Forensics

Martin Potthast, Francisco Rangel, Michael Tschuggnall, Efstathios Stamatatos, Paolo Rosso, and Benno Stein

Author Identification

Style Breach Detection

- Given a document, find the positions where the authorship, i.e., the style, changes
- Closely related to *intrinsic plagiarism detection* and general *text segmentation* problems
- Data: 289 English documents, created from Webis-TRC-12
- baseline placing random borders such that fragments are equally distributed

Results	WinP	WinR	WinF	WindowDiff
Karaś *et al.*	0.315	0.586	**0.323**	0.546
BASELINE-eq	0.337	**0.645**	0.289	0.647
Khan	**0.399**	0.487	0.289	**0.480**
Safin *et al.*	0.371	0.543	0.277	0.529

Author Clustering

- Given a set of short (paragraph-length) single-author documents, group them by authorship.
- *Complete clustering*: the number of distinct authors is not given.
- *Authorship-link ranking*: extract pairs of documents by the same author and rank them by an estimated confidence score.
- Data: 180 clustering instances in 3 languages (English, Dutch, Greek) and 2 genres (articles, reviews).
- 6 submissions and 4 baseline models.

Top results	B^3 F	MAP
Gómez-Adorno et al.	**0.573**	**0.456**
BASELINE-PAN16	0.487	0.443

Author Profiling

Objective Given tweets writen in a language (e.g. English), identify their author's gender and the language variety used (e.g. British).

Dataset 500 authors per variety and gender, 300 for training, 200 for test, with 100 tweets per author.

Arabic	English	Spanish	Portuguese
Egypt	Australia	Argentina	Brazil
Gulf	Canada	Chile	Portugal
Levantine	Great Britain	Colombia	
Maghrebi	Ireland	Mexico	
	New Zealand	Peru	
	United States	Spain	
		Venezuela	
4,000	6,000	7,000	2,000

Approaches

- 22 participants from 19 countries.
- *Features*: Classical (n-grams, content&style-based) vs. Embeddings.
- *Algorithms*: Classical (Logistic regression, SVM, Naive Bayes, ...) vs. Deep learning (Recurrent Neural Networks, Convolutional Neural Networks, ...).

Best results

Language	Joint	Gender	Variety
Arabic	0.6831	0.8031	0.8313
English	0.7429	0.8233	0.8988
Portuguese	0.8575	0.8700	0.9838
Spanish	0.8036	0.8321	0.9621

Some conclusions

- Deep learning approaches have not obtained the best results.
- The best results have been obtained in Portuguese for all tasks.
- Most difficult varieties per language: Gulf (Arabic), Australia (English), Portugal (Portuguese), Peru (Spanish).

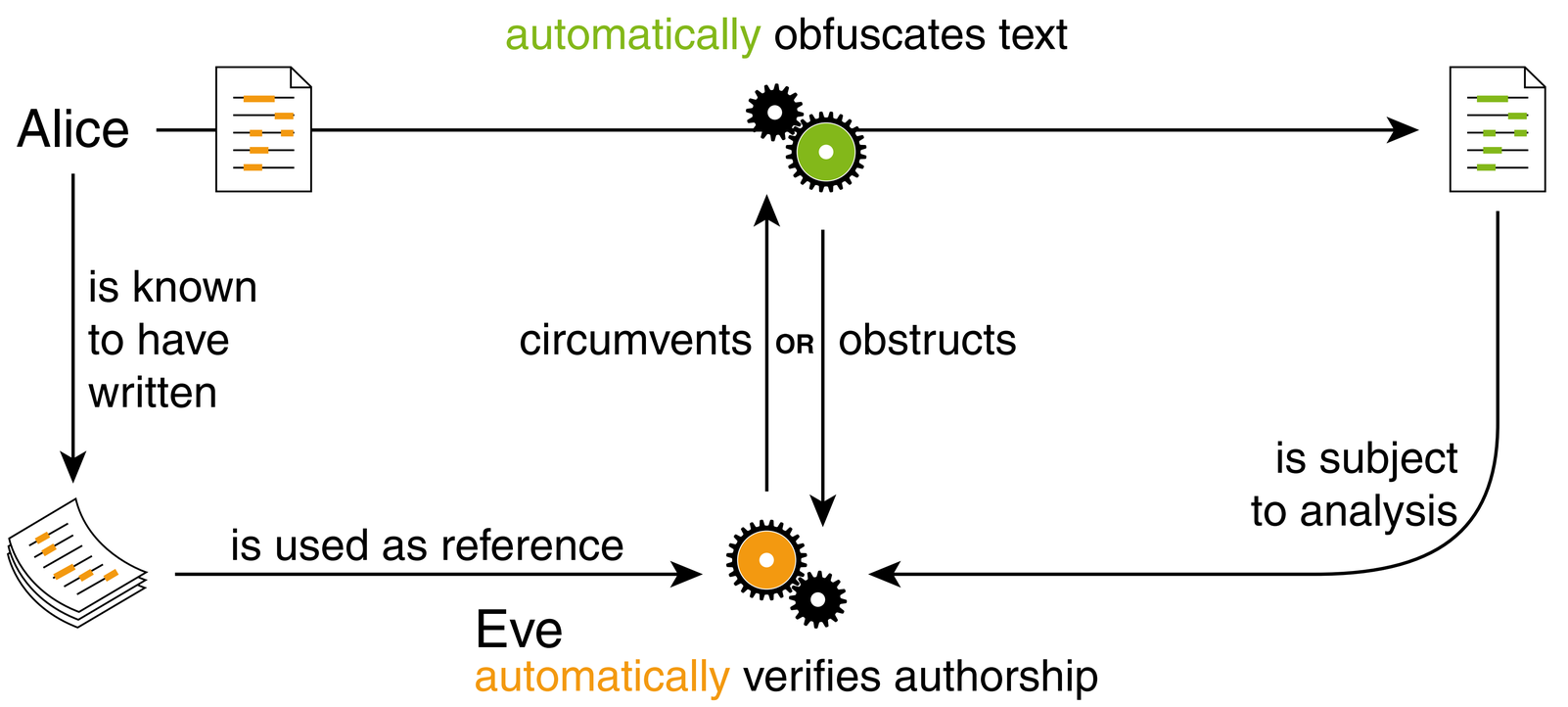
Author Obfuscation

Authorship Verification

Given two documents, decide whether both have been written by the same author.

Author Masking

Given two documents from the same author, paraphrase the designated one such that an authorship verification will fail.



- 5 submitted obfuscators PAN'16/17
- Against 44 verifiers from PAN'13-15
- Key insight: about 50% of true positive decisions can be flipped to false negative