



Overview of the Cross-Domain Authorship Attribution Task at PAN 2019

Mike Kestemont, Efsthios Stamatatos, Enrique Manjavacas, Walter Daelemans, Martin Potthast, and Benno Stein
Lugano, 11 Sept 2019

Task

- Author identification: given anonymous document, determine who wrote it on basis of writing style
 - (Closed-set) Attribution: traditional text categorisation perspective
 - Feasible within single domain (genre) with enough training data
- Characteristics PAN 2019: more realistic and difficult setup
 - Open-set attribution: back-off possibility, for <UNK> authors
 - Cross-domain attribution: train and test don't overlap in domains

Fan fiction



- “Transformative” literature by non-professional authors (“fans”), following a specific example
- Fastest growing form of writing; massive availability
- High-quality metadata (e.g. labels for “fandoms” as “Harry Potter”)
- Unmediated: self-published, with no moderation

Problems

r = adversary ratio = proportion of adversary authors (<UNK>)

	Language	# Problems	$ A $	$ K_a $	$ U $	r	\bar{l}
Development	English	5	9	7	137-561	0.2 - 1.0	804
	French	5	9	7	38-430	0.2 - 1.0	790
	Italian	5	9	7	46-196	0.2 - 1.0	814
	Spanish	5	9	7	112-450	0.2 - 1.0	846
Evaluation	English	5	9	7	98-180	1.0	817
	French	5	9	7	48-290	1.0	790
	Italian	5	9	7	34-302	1.0	821
	Spanish	5	9	7	172-588	1.0	838

All obtained from archiveofourown.org

Evaluation

- Open-set, macro-averaged F1 score (imbalance of test labels)
- Three baselines (based on character n-grams)
 1. BASELINE-SVM: $Pr1 - Pr2 < 0.1 \rightarrow \langle \text{UNK} \rangle$
 2. BASELINE-COMPRESSOR: $(S_1 - S_2) / S_1 < 0.01 \rightarrow \langle \text{UNK} \rangle$
 3. BASELINE-IMPOSTERS: $p > .10 \rightarrow \langle \text{UNK} \rangle$
- Posthoc majority baseline: accept majority vote, else $\langle \text{UNK} \rangle$

Background corpus

- Novelty this year (but hardly used)
- 5,000 documents per language
- Known topics (!), unknown authors (no overlap with problems)
- Useful for:
 - imposter verification (now on *ad hoc* corpora)
 - language/topic model pretraining

Main results

$r = 1.0$ (equal # of adversaries) | macro av. per language

Submission	Macro-Precision	Macro-Recall	Macro-F1	Runtime
Muttenthaler et al.	0.716	0.742	0.690	00:33:17
MAJORITY	0.748	0.708	0.686	
Bacciu et al.	0.688	0.768	0.680	01:06:08
Custodio & Paraboni	0.664	0.717	0.65	01:21:13
Bartelds & de Vries	0.657	0.719	0.644	11:19:32
Rodríguez et al.	0.651	0.713	0.642	01:59:17
Isbister	0.629	0.706	0.622	01:05:32
Johansson	0.593	0.734	0.616	01:05:30
Basile	0.616	0.692	0.613	00:17:08
van Halteren	0.590	0.734	0.598	37:05:47
Rahgouy et al.	0.601	0.633	0.580	02:52:03
Gagala	0.689	0.593	0.576	08:22:33
BASELINE-SVM	0.552	0.635	0.545	
BASELINE-COMPRESSOR	0.561	0.629	0.533	
BASELINE-IMPOSTERS	0.428	0.580	0.395	
Kipnis	0.270	0.409	0.259	20:20:21

Vary adversary ratio

Lower r yields higher performance

Submission	r				
	100%	80%	60%	40%	20%
Muttenthaler et al.	0.690	0.709	0.727	0.746	0.773
Bacciu et al.	0.680	0.701	0.724	0.749	0.777
Custodio & Paraboni	0.650	0.666	0.686	0.704	0.728
Bartelds & de Vries	0.644	0.663	0.683	0.708	0.736
Rodríguez et al.	0.642	0.664	0.684	0.704	0.733
Isbister	0.622	0.642	0.664	0.685	0.716
Johansson	0.616	0.641	0.666	0.700	0.735
Basile	0.613	0.633	0.654	0.675	0.706
van Halteren	0.598	0.622	0.645	0.672	0.701
Rahgouy et al.	0.580	0.599	0.619	0.642	0.664
Gagala	0.576	0.586	0.597	0.610	0.624
BASELINE-SVM	0.545	0.563	0.585	0.611	0.642
BASELINE-COMPRESSOR	0.533	0.548	0.569	0.592	0.620
BASELINE-IMPOSTERS	0.395	0.409	0.429	0.453	0.484
Kipnis	0.259	0.270	0.285	0.302	0.324

Stable across languages

Move to Asian languages in future?

Submission	English	French	Italian	Spanish
Muttenthaler et al.	0.665	0.705	0.717	0.673
Bacciu et al.	0.638	0.689	0.715	0.679
Custodio & Paraboni	0.587	0.686	0.682	0.647
Bartelds & de Vries	0.558	0.687	0.700	0.629
Rodríguez et al.	0.597	0.624	0.696	0.651
Isbister	0.529	0.644	0.691	0.623
Johansson	0.613	0.593	0.655	0.602
Basile	0.555	0.628	0.656	0.613
van Halteren	0.532	0.554	0.653	0.652
Rahgouy et al.	0.550	0.583	0.595	0.592
Gagala	0.554	0.564	0.566	0.619
BASELINE-SVM	0.490	0.548	0.566	0.577
BASELINE-COMPRESSOR	0.493	0.595	0.580	0.464
BASELINE-IMPOSTERS	0.359	0.409	0.410	0.400
Kipnis	0.301	0.232	0.285	0.220

Limited diversity

Participant	Features	Weighting	Feature transformation/selection	Parameter tuning	Classifier	Open-set criterion	Use imposters data	Language-dependent resources
Bacciu et al.	n-grams (char, word, POS, stem, distortion)	tf-idf	NA	per language	Ensemble (SVM)	Reject	No	stemming, POS
Bartelds and de Vries	n-grams (char, token, POS, punctuation, ...)	tf-idf	NA	global	SVM	Reject	No	POS, syntactic parse
Basile	n-grams (char and word)	tf-idf	NA	global	SVM	Reject	No	None
Custodio et al.	n-grams (char, word, distortion), ...	tf-idf	PCA	global	Ensemble (LR)	Reject	No	None
Gagala	n-grams (char, word)	tf-idf	PCA	global	Imposters (LR)	Verification	Yes	None
Isbister	n-grams (char, word), word and sentence ...	tf-idf	NA	global	SVM	Reject	No	None
Johansson	n-grams (char, word, POS, distortion), word ...	tf-idf	NA	global	SVM	Reject	No	POS
Muttenthaler et al.	n-grams (char, word, distortion, punctuation)	tf-idf	SVD	global	Ensemble (SVM)	Reject	No	None
Rahgouy et al.	n-grams (char and word), word embeddings	tf-idf and tf	NA	global	Ensemble (SVM)	Reject	No	stemming
Rodríguez et al.	n-grams (char, typed, punctuation, word)	tf-idf	NA	global	Ensemble (SVM)	Reject	No	None
Van Halteren	n-grams (char, token, syntactic)	z-score	NA	per language	Ensemble (distance-based and SVR)	Reject	No	POS, syntactic parse
baseline-SVM	n-grams (char)	tf	NA	global	SVM	Reject	No	None
baseline-Compressor	char sequences	none	NA	global	PPM	Reject	No	None
baseline-Imposters	n-grams (char)	tf	NA	global	distance-based	Verification	Yes	None

Conclusions: *quo vadis?*

- Limited diversity of approaches: shared task should stimulate diversity, rather than convergence
- No neural networks: TIRA not ready, but not very promising direction?
- Cross-domain information not exploited
- Move to verification for PAN2020

#fanfics

Because the story's not over until we say it is.

See you at PAN2020