

Time-Aware Authorship Attribution for Short Text Streams

Hosein Azarbyonad, Mostafa Dehghani, Maarten Marx, and Jaap Kamps
University of Amsterdam

Motivation

Most cybercrimes are done by exchanging short texts.



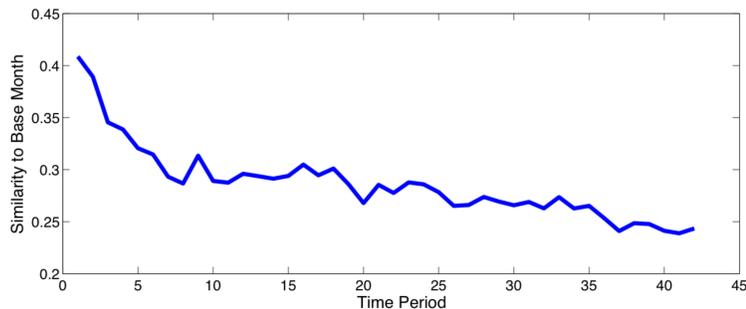
Cybercriminals tend to use an anonymous identity on the Internet based communication systems.

Authorship Attribution:

- Assign a given text of unknown authorship to one of candidate authors for whom text samples of undisputed authorship are available.

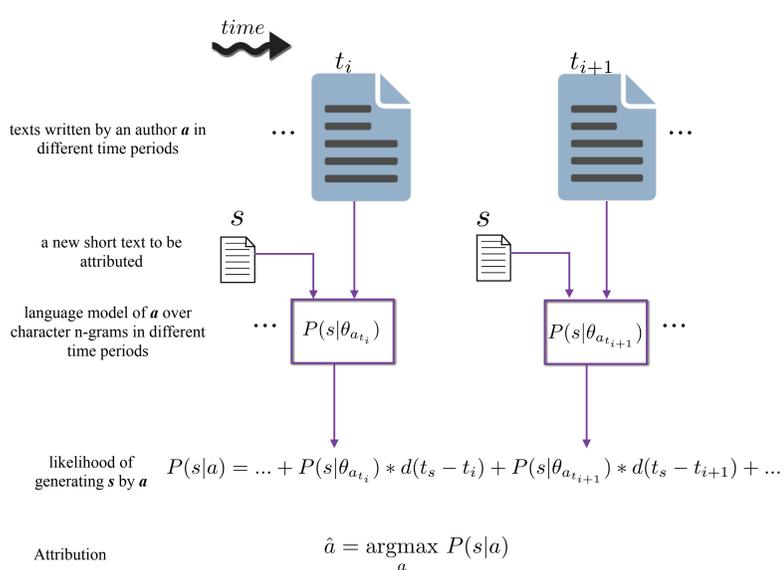
Challenges:

- Shortness of texts
- Dynamicity of authors



Vocabulary usage changes of Twitter users over time. The first two months of the users' activity in Twitter are considered as start period. Each following month is considered as a time period. The x -axis shows the time periods and y -axis shows the averaged cosine similarity of the contents generated by the users at each time period with the content generated by them in the base time period.

Time-Aware Authorship Attribution



Decay Functions:

- General Decay Factor:** Estimated based on the plot of average similarity to the start period. A line is fitted to this plot using linear regression:

$$d(t) = \frac{1}{Z}(at + b)$$

- Specific Decay Factor:** Estimated using the plot of similarity to start period for each user:

$$d_a(t) = \frac{1}{Z_a}(a_a t + b_a)$$

Experiments

Data Sets

- Twitter
 - 133 users: avg. number of tweets per user=1820, timeline=[2010-01,2014-10]
- Enron email dataset
 - 15 prolific users: avg. number of tweets per user=3200, timeline=[1998-01,2002-09]
- Each month is considered a time period
- Train data: 90% oldest texts are used for constructing the language models
- Test data: 10% of newest texts written by each author

Experiment 1: Drift in word usage over time

- Research Question 1:** Does the writing style of authors of short text change over time? And if so, do they change their writing styles by the same rate?

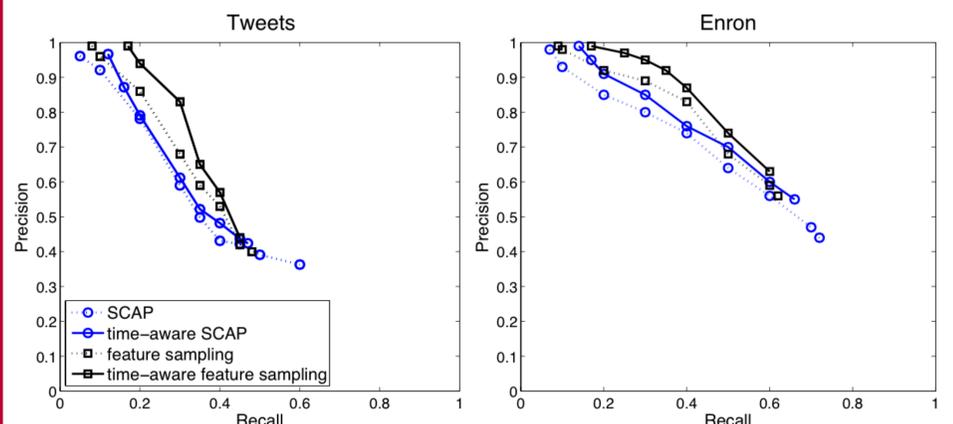
Dataset	max	min	average	median	std
Enron(N=15)	0.09	-0.1	-0.007	-0.006	0.161
Tweets(N=133)	0.06	-0.07	-0.001	-0.002	0.07

The statistics of slope a_a of the specific decay function for different authors on Enron and Tweet dataset.

- Outcome 1:** The difference of changes of authors' vocabulary usages is considerable.

Experiment 2: Time-aware authorship attribution

- Research Question 2:** How effective is the time-aware authorship attribution approach compared to time-unaware authorship attribution approaches?



Precision-Recall curves of time-aware and time-unaware authorship attribution methods on tweets and Enron datasets.

- Outcome 2:** Time-aware authorship attribution approaches successfully outperform time-unaware approaches.

Dataset	Time-unaware	Time-aware(general)	Time-aware(specific)
Enron	0.87	0.88	0.94 (8%) [▲]
Tweets	0.69	0.71	0.80 (15%) [▲]

Precision of feature sampling method on tweets and Enron dataset at Recall point of 0.3. (▲ indicates the significance using t-test, one-tailed, p -value < 0.05)

- Outcome 3:** Using the specific decay function gives significantly better performance compared to using the general decay function.

Conclusions

- Authors change their writing styles and vocabulary usage over time.
- The vocabulary usage change rate is different for different authors.
- Considering the dynamicity of authors improves the accuracy of authorship attribution.
- For different authors, different decay functions based on their vocabulary usage change rates should be considered in authorship attribution process.