Detecting a Change of Style Using Text Statistics

Kamil Safin   Aleksandr Ogaltsov

Antiplagiat Company
Moscow Institute of Physics and Technology
Higher School of Economics

PAN'18 competition

Tasks

- Author identification task.
  - Document written by one author or not.
  - Binary classification task.
- Author profiling task.
- Author obfuscation task.

Author Identification

Given a document, who wrote it?

One subtask focuses on **cross-domain authorship attribution** applied in fanfiction and another subtask focuses on **style change detection**.

Learn more »

Author Profiling

Given a document, what're its author's traits?

This task focuses on **gender**, whereas **text and image** may be used as information sources of tweets in English, Spanish and Arabic.

Learn more »

Author Obfuscation

Given a document, hide its author.

This task works against identification and profiling by automatically paraphrasing a text to obfuscate its author's style. The tasks offered are **author masking** and **obfuscation evaluation**.

Learn more »

Style change detection

Given a document, determine whether it contains style changes or not.

- **Yes** — the document contains at least one style change.
- **No** — the document has no style changes.

Data

The data corpus consists of user posts from various sites of the StackExchange network.

Site	Training				Validation				Test			
	Problems	Authors			Problems	Authors			Problems	Authors		
		1	2	3		1	2	3		1	2	3
bicycles	160	80	47	33	82	41	28	13	70	35	27	8
christianity	358	179	107	72	176	88	48	40	172	86	45	41
gaming	178	89	47	42	86	43	23	20	78	39	21	18
history	354	177	104	73	178	89	54	35	170	85	46	39
islam	166	83	49	34	86	43	31	12	72	36	20	16
linguistics	144	72	46	26	72	36	22	14	64	32	12	20
meta	196	98	56	42	94	47	30	17	90	45	30	15
parenting	178	89	54	35	92	46	32	14	78	39	27	12
philosophy	468	234	146	88	232	116	63	53	224	112	65	47
poker	100	50	35	15	48	24	14	10	42	21	13	8
politics	204	102	57	45	102	51	34	17	90	45	22	23
project man.	104	52	24	28	50	25	12	13	44	22	14	8
sports	102	51	34	17	54	27	20	7	40	20	12	8
stackoverflow	112	56	23	33	60	30	16	14	48	24	12	12
writers	156	78	43	35	80	40	25	15	70	35	18	17
Σ	2980	1490	872	618	1492	746	452	294	1352	676	384	292

Metaclassifier

Components

- Statistical Classifier — p_s.
- Hashing Classifier — p_h.
- Counting Classifier — p_c.

Final Score

$$\text{score}(d) = \alpha_s p_s + \alpha_h p_h + \alpha_c p_c,$$

α_j — weights of each classifier, $\sum \alpha_i = 1$.

Classification

$$\text{score}(d) > \delta \Rightarrow d \text{ has change of style},$$

d — document, δ — classification threshold.

Metaclassifier

Quality criteria

Accuracy as measure of quality:

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}.$$

Statistical Classifier

- Collector of statistical features, such as:
    - number of sentences,
    - unique words fraction,
    - text length,
    - punctuation symbols fraction,
    - letter symbols fraction, etc.

- 19-dimensional feature space.

- Random Forest for final proba.

Metaclassifier

Hashing Classifier

- Hashing function to build term frequency counts.
- 3000-dimensional representation space.
- Random Forest for final proba.

Counting Classifier

- Word n-grams counts form 1 to 6.
- High-dimensional (3M) representation of text.
- Logistic Regression for final proba.

Parameters Tuning

- Tune $\alpha_s, \alpha_h, \alpha_c$ and threshold δ;
- $\alpha_s, \alpha_h, \alpha_c$ shows the importance of corresponding classifier;
- Optimal: $\alpha_s = 0.4$, $\alpha_h = 0.2$, $\alpha_c = 0.4, \delta = 0.55$.
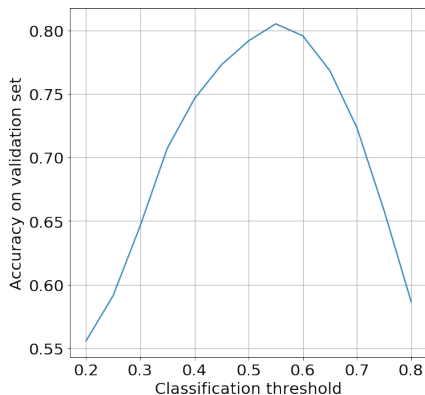
Results

The proposed model was tested on PAN'18 data set. The results of its performance are shown below.

	Validation	Test
Accuracy	0.805	0.803

Comparison with other participants is shown below.

Submission	Accuracy	Runtime
Zlatkova et al.	0.893	01:35
Hosseinia and Mukherjee	0.825	10:12
Safin and Ogaltsov	**0.805**	**00:05**
Khan	0.643	00:01
Schaetti	0.621	00:03

Q & A
Tnank you for attention!