

Gender Profiling in Social Network

Постановка задачи

Тема: Робастные (несмещенные) методы корпусного исследования социолингвистической вариативности [на примере языка соцсетей]

Задача: Одножанровое и многожанровое гендерное профилирование авторов по набору текстов соцсетей

При составлении *Обзора работ* использовались следующие источники:

- [Материалы PAN CLEF](#) за 2016 [t1], 2017 [t2], 2019 годы [t3]
- Материалы RusProfiling PAN at FIRE Track за 2017 год [d8]
- Отдельные исследования на тему гендерного профилирования в социальных сетях как для русского, так и других языков ([x10], [x12], [r1], [r4], [r6])
- Другие материалы

В рассматриваемых работах использовались следующие датасеты:

Датасеты

PAN '16 [d1] [d2] d[3]

тексты из Twitter на Английском, Испанском (PAN-AP14 Twitter corpus [d1]) и Голландском языках (TwiSty [d3]).

Для тестирования - отзывы об отелях на Голландском языке (CSI corpus [d2]); блоги на Английском и Испанском языках [d1]

BT-AP-19 [d4]

тексты из Twitter на Арабском языке

from Heaven BBS [d5]

тексты сообщений из BBS на Турецком языке

PAN '17 [d6]

тексты из Twitter на Английском, Испанском, Арабском и Португальском языках

- Арабский: Египет, Персидский залив, Левант, Магриб
- Английский: Австралия, Канада, Великобритания, Ирландия, Новая Зеландия, США.
- Португальский: Бразилия, Португалия.
- Испанский: Аргентина, Чили, Колумбия, Мексика, Перу, Испания, Венесуэла

PAN '19 [d7]

тексты из Twitter на Английском, Испанском языках

PAN FIRE '17 [d8]

тексты из Twitter, Facebook, отзывы, имитации письма противоположным полом на Русском языке

RusPersonality [d9]

Дневниковые записи, письма, эссе на Русском языке

LiveJournal [d10]

Блоги 1290 авторов LiveJournal на Русском языке

RUEN-AP-17 [d13]

Посты и комментарии пользователей Facebook на Английском и Roman-Urdu

SMS-AP-18 [d14]

SMS пользователей на Английском и Roman-Urdu

Обзор работ

Ниже приведены таблицы с результатами исследований следующих задач:

- Определение пола автора в одножанровых текстах соцсетей (все языки, кроме русского)
- Определение пола автора в одножанровых русскоязычных текстах соцсетей
- Определение пола автора в многожанровых русскоязычных текстах соцсетей
- Определение пола автора в многожанровых текстах соцсетей (все языки, кроме русского)

Определение пола автора в одножанровых текстах соцсетей (все языки, кроме русского)

В *Таблице 1* приведены данные по гендерному профилированию в одножанровых текстах соцсетей для различных языков (кроме русского), проводившемся в разные годы, в которых достигнуты высшие показатели.

По столбцам приведены Авторы и ссылка на работу, год публикации, корпус, язык корпуса, используемые признаки, метод исследования, результат.

Лучший результат у *Ashraf and Nie [x10] (2020)*, который использовал 100-dimension Twitter-based GloVe pre-trained корпус и Bi-LSTM. Высокий итоговый показатель, видимо, обусловлен хорошим эмбедингом представления слов.

В работе [x6] для английского языка использовался BERT. Результат оказался сопоставим с другими подходами, рассматриваемыми авторами и, по их признанию, модель с BERT требует более тонкой настройки для получения лучшего результата.

Согласно результатам PAN '19, традиционные подходы получили более высокую точность, чем методы глубокого обучения. Команды, показавшие четыре лучших результата, использовали комбинацию n-gramm и SVM, а лучшая команда [x9], применившая модель глубокого обучения, оказалась на одиннадцатом месте.

В целом, при одинаковых датасетах, команды добивались наивысших показателей , в основном, работая со стилометрическими признаками (word, character n-gramms, POS, etc.), в сочетании с SVM.

TABLE I. One-genre Gender Detection (languages w/o Russian)

| Work | Publishing Year | Corpus | Text languages | Features | Method used | Result |
|----------------------|-----------------|---------------|-----------------------|---|---------------------|--------|
| Ashraf and Nie [x10] | 2020 | BT-AP-19 [d4] | English Roman-Urdu | GloVe | Bi-LSTM | 0.88 |
| Valencia et al. [x8] | 2019 | PAN '19 [d7] | English | char(3-4), word(1-3) n-grams, tf-idf | Logistic regression | 0.8432 |
| Gishamer [x11] | 2019 | PAN '19 [d7] | English | char(3-5); word(1-3); POS; POS(2-3) - grams | SVM | 0.84 |
| Basile et al. [x3] | 2017 | PAN '17 [d6] | Arabic | char(3-5); word1- 2) - grams; [tf_ = 1 + log(tf)] - idf; | SVM | 0.8006 |
| | | | English | | | 0.8233 |
| | | | Portuguese | | | 0.8450 |
| | | | Spanish | | | 0.8321 |
| Joo & Hwang [x6] | 2019 | PAN '19 [d7] | English | lexical, tf-idf, Doc2Vec(DBOW+DM 1+2+3-gram), used words | features + BERT | 0.8360 |
| Pizarro [x4] | 2019 | PAN '19 [d7] | English | char(1, 3), word(1, 3) n-grams, tf-idf | SVM | 0.8356 |
| | | | | | | |

| | | | | | | |
|---------------------------|------|----------------------|------------|--|----------------------------|--------|
| | | | Spanish | char(3, 5), word(1, 3) n-grams, tf-idf | | 0.8172 |
| | | | Spanish | char(3-5); word1-2);POS(2-3) -grams; | | 0.712 |
| Tayfun et al [x12] | 2006 | from Heaven BBS [d5] | Turkish | stylistic features, CHI square | Naive Bayesian Style-based | 0.819 |
| Martinc et al. [x1] | 2017 | PAN '17 [d6] | Arabic | word(1, 2), char(4), suffix(4), punctuation(3) n-grams | Logistic regression | 0.8031 |
| | | | English | | | 0.8071 |
| | | | Portuguese | | | 0.8600 |
| | | | Spanish | | | 0.8193 |
| Srinivasarao & Manu [x5] | 2019 | PAN '19 [d7] | English | char(3, 4), word(1, 3) n-grams, tf-idf | SVM | 0.8398 |
| | | | Spanish | char(3, 4), word(1, 2) n-grams, tf-idf | | 0.7967 |
| Goubin et al. [x7] | 2019 | PAN '19 [d7] | English | BOW, tf-idf | LSVC + stat calibration | 0.8333 |
| | | | Spanish | | | 0.7917 |
| Miura et al. [x2] | 2017 | PAN '17 [d6] | Arabic | word & character embedding | RNN, CNN | 0.7644 |
| | | | English | | | 0.8046 |
| | | | Portuguese | | | 0.8700 |
| | | | Spanish | | | 0.8118 |
| Polignano & de Pinto [x9] | 2019 | PAN '19 [d7] | English | | CNN, Word2Vec | 0.7973 |
| | | | Spanish | | | 0.7417 |
| | | | Spanish | | | 0.7539 |

Определение пола автора в одножанровых русскоязычных текстах соцсетей

В Таблице 2 приведены данные по гендерному профилированию в одножанровых русскоязычных текстах соцсетей, проводившемуся в разные годы, в которых достигнуты высшие показатели.

По столбцам приведены Авторы и ссылка на работу, год публикации, корпус, используемые признаки, метод исследования, результат.

Наилучший результат был достигнут в 2013 году [r6]. Однако, с учетом того, что был использован тот же метод, что и в работах позже (PAN FIRE '17), достижение, возможно, обусловлено спецификой датасета и тестовых примеров, на которых происходила оценка работы модели.

Среди работ, где использовался датасет PAN FIRE '17 [d8], выделяется [r4], сделанная позже, в 2019 г. В ней были проведены эксперименты с различными моделями, в том числе глубокого обучения. На первом месте оказалась Gradient Boosting модель с символьным n-gramm представлением.

В то же время, в работе [r1] 2020 года, результат оказался выше при использовании модели с GRU, CVAE. Стоит отметить, что в работе не упомянуто сравнение с другими моделями на используемом датасете.

TABLE II. One-genre Gender Detection (Russian)

| Work | Publishing Year | Corpus | Features | Method used | Result |
|--------------------------|-----------------|---|---|-------------------|--------|
| Korshunov [r6] | 2013 | self-made | word (3-grams) | SVM | 0.86 |
| Sboev et al [r4] | 2019 | RusProfilihg [d11] PAN FIRE '17 [d8] | char n-grams | Gradient Boosting | 0.79 |
| Сбоев и др. [r1] | 2020 | LiveJournal [d10] | | GRU, CVAE | 0.76 |
| Markov et al - CIC3 [r2] | 2017 | PAN FIRE '17 [d8] | statistical | | 0.6825 |
| LDR [d12] | 2017 | PAN FIRE '17 [d8] | probability distribution of occurrence of tdoc's words in | | 0.6759 |

| | | | | | |
|--------------------------|------|-------------------|------------------------------------|---------------|--------|
| | | | the different classes. | | |
| Markov et al - CIC2 [r2] | 2017 | PAN FIRE '17 [d8] | BOW, word (suffix 3-grams), tf-idf | SVM | 0.6650 |
| Bhargava et al [r3] | 2017 | PAN FIRE '17 [d8] | POS, rule-based classification | LSTM, Bi-LSTM | 0.6525 |
| Markov et al - CIC1 [r2] | 2017 | PAN FIRE '17 [d8] | POS combination, tf-idf | SVM | 0.6525 |

Определение пола автора в многожанровых русскоязычных текстах соцсетей

В Таблице 3 приведены высшие показатели по гендерному профилированию в многожанровых русскоязычных текстах соцсетей, на основе результатов RusProfiling PAN at FIRE Track, проводившегося в 2017 году.

По столбцам приведены Авторы и ссылка на работу, год публикации, обучающий корпус, используемые признаки, метод исследования, тестовые корпуса (эссе, Facebook, отзывы и тексты, в которых авторы имитировали другой пол, а пользователи меняют свой идиостиль).

В каждой из четырех тестовых соревнованиях, выбранный в качестве baseline, метод LDR [d12], оказывался в лидерах. Ключевое понятие метода - вес, представляющий вероятность того, что термин принадлежит к одной из категорий. Распределение весов для рассматриваемого документа должно быть ближе к весам соответствующей категории.

Среди эссе лучшего результат показан в работе [w5], где применялась комбинация rule-based classification и методов глубокого обучения.

Неожиданными оказались довольно высокие результаты на тестовой выборке из *Facebook*, которые превосходили все остальные категории, включая *Twitter*.

По гипотезе авторов обзора [d8], это обусловлено тем, что в *Facebook* сообщения длиннее и грамматически богаче, с меньшим количеством синтаксических ошибок и опечаток по сравнению с *Twitter*.

По результатам, модели с традиционным подходом, оказались лучше, чем модели, использующие методы глубокого обучения.

TABLE III. Cross-genre Gender Detection (Russian)

| | | | | | |
|--|--|--|--|--|--|
| | | | | | |
|--|--|--|--|--|--|

| Work | Publishing Year | Corpus | Features | Method used | Test Corpus | Result |
|--------------------------|-----------------|--|---|---------------|-------------|--------|
| LDR [d12] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | stylometric analysis | | Essays | 0.8141 |
| Bhargava et al [r3] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | POS, rule-based classification | LSTM, Bi-LSTM | Essays | 0.7838 |
| Vinayan et al [r5] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | exotic stat (average word length, URL usage, etc), tf-idf | SVM | Essays | 0.6811 |
| Markov et al - CIC2 [r2] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | BOW, word (suffix 3-grams), tf-idf | SVM | Facebook | 0.9342 |
| Markov et al - CIC1 [r2] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | POS combination, tf-idf | SVM | Facebook | 0.9211 |
| Markov et al - CIC5 [r2] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | word(1), char(2-3) n-grams, tf-idf | SVM | Facebook | 0.8991 |
| Markov et al - CIC4 [r2] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | word(1, 3), char(2-4) n-grams, tf-idf | SVM | Facebook | 0.8860 |
| Bhargava | 2017 | PAN | POS, rule- | LSTM, Bi- | Facebook | 0.8728 |

| | | | | | | |
|--------------------------|------|--|--------------------------------|-------------------|------------------|--------|
| et al [r3] | | FIRE '17 [d8] (Twitter - training) | based classification | LSTM | | |
| LDR [d12] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | stylometric analysis | | Facebook | 0.8596 |
| Sboev et al [r4] | 2019 | PAN FIRE '17 [d8] (Twitter - training) | char n-grams | Gradient Boosting | Reviews | 0.79 |
| LDR [d12] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | stylometric analysis | | Reviews | 0.72 |
| Markov et al - CIC3 [r2] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | statistical | | Reviews | 0.6186 |
| Markov et al - CIC1 [r2] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | POS combination, tf-idf | SVM | Reviews | 0.5979 |
| Bhargava et al [r3] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | POS, rule-based classification | LSTM, Bi-LSTM | Reviews | 0.5786 |
| Bhargava et al [r3] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | POS, rule-based classification | LSTM, Bi-LSTM | Gender imitation | 0.6596 |

| | | | | | | |
|-----------|------|--|----------------------|--|------------------|--------|
| LDR [d12] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | stylometric analysis | | Gender imitation | 0.6383 |
|-----------|------|--|----------------------|--|------------------|--------|

Определение пола автора в многожанровых текстах соцсетей (все языки, кроме русского)

В Таблице 4 приведены высшие показатели по гендерному профилированию в многожанровых текстах соцсетей (все языки, кроме русского), на основе результатов [PAN CLEF](#), проводившегося в 2016 году и работе [x10].

По столбцам приведены Авторы и ссылка на работу, год публикации, обучающий корпус, используемые признаки, метод исследования, тестовые корпуса (соцсети, блоги, отзывы).

Как и в исследованиях одножанровых текстов, лучший результат у *Ashraf and Nie* [x10] (2020), который использовал 100-dimension Twitter-based GloVe pre-trained корпус и Bi-LSTM в тестировании текстов на *Facebook*, и CNN для SMS-сообщений.

В PAN CLEF 2016 года большинство участников рассматривали стилистические особенности текстов (частота использование служебных слов, сленг, использование заглавных букв и т.п.) в сочетании с POS, n-grams, с не особо впечатляющими результатами.

TABLE IV. Cross-genre Gender Detection (languages w/o Russian)

| Work | Publishing Year | Corpus | Features | Method used | Test Corpus | Result |
|----------------------|-----------------|------------------------------------|----------|-------------|--|--------|
| Ashraf and Nie [x10] | 2020 | BT-AP-19 [d4] (Twitter - training) | GloVe | Bi-LSTM | RUEN-AP-17 [d13], Facebook (English, Roman-Urdu) | 0.833 |
| Ashraf and Nie [x10] | 2020 | BT-AP-19 [d4] (Twitter - training) | GloVe | CNN | SMS-AP-18 [d14], SMS (English, Roman-Urdu) | 0.815 |
| Busger et | 2016 | PAN '16 | word(1- | SVM | social | 0.5575 |

| | | | | | | |
|--------------------------------|------|--|---|------------------------|------------------------------|--------|
| al. [s1] | | [d1] [d2] d[3] (Twitter - training) | 3),char(2-5) n-grams, POS | | media (English) | |
| Waser | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | not presented | not presented | social media (Spanish) | 0.7031 |
| Roman- Gomez | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | not presented | not presented | social media (Dutch) | 0.6200 |
| Agrawal & Gonçalves [s2] | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | tf-idf | SVM, NB | social media (English) | 0.5431 |
| Modaresi et al. [s3] | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | word(1- 2),char(4) n- grams, punctuation | Logistic regression | social media (Spanish) | 0.6406 |
| Waser | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | not presented | not presented | social media (Dutch) | 0.6000 |
| Modaresi et al. [s3] | 2016 | PAN '16 [d1] [d2] d[3] | word(1- 2),char(4) n- grams, punctuation | Logistic regression | blogs (English) | 0.7564 |

| | | | | | | |
|------------------------|------|---|---|----------------|-----------------|--------|
| | | (Twitter - training) | | | | |
| Deneva | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | not presented | not presented | blogs (Spanish) | 0.7321 |
| Bilan et al. [s4] | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | word(1-3) n-grams, tf-idf, CHI square, POS, stylistic | SVM(LinearSVC) | blogs (English) | 0.7436 |
| Busger et al. [s1] | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | word(1-3),char(2-5) n-grams, POS | SVM | blogs (Spanish) | 0.7143 |
| Deneva | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | not presented | not presented | reviews (Dutch) | 0.6180 |
| Bayot & Gonçalves [s5] | 2016 | PAN '16 [d1] [d2] d[3] (Twitter - training) | CBOW | Word2Vec, SVM | reviews (Dutch) | 0.5680 |

Подходы

При построении моделей гендерного профилирования одно- и многожанровых текстов на различных языках из соцсетей, авторами исследовались подходы, которые можно свести к следующим:

1. Методы стилометрии, цель которых - уловить стиль письма автора с использованием различных статистических характеристик

a. лексические

использование определенных букв, слов, смайлов, символьные n-граммы, частота их использования и т.п.

b. синтаксические

использование пунктуации, POS, частоты знаков препинания, и т.п.

c. структурные

то, как автор структурирует текст или использует определенные шаблоны (приветствия, подписи и т.п.)

2. Методы, основанные на содержании, которые предназначены для идентификации профилей автора на основе содержания текста (BOW, n-граммы слов, сленговые слова и т.п.)

a. *tf-idf*

Цель - оценить важность слова в контексте документа, являющегося частью коллекции документов или корпуса. Такая стратегия была предложена многочисленными стилометрическими исследованиями [m1], [m2], [m3].

Однако, после проверки значений *tf-idf*, заметных различий между мужчинами и женщинами не выявилось. Т.е., напрямую использовать данную характеристику для того, чтобы различать гендер автора, не получится.

Тем не менее, важность метода - выделение специфических для соцсетей признаков с высоким рейтингом упоминаний - например, для Twitter это (@), гиперссылки (urllink), ретвиты (rt) и хэштеги (#). Для случая персонального общения - частое употребление личных местоимений (например, я, оно, ты), а также некоторые знаки препинания (.,:) и функциональные термины (the, to, of, is, and, in) (для английского языка).

b. Выделить специфические слова для каждой из категорий

Т.е. когда термины часто используются в одной категории и редко в другой. Тогда разность в вероятности появления должна обнаружить все, или, хотя бы, некоторые из них.

$$D_{tf}(t_i) = p(t_i, M) - p(t_i, F) = \frac{tf_{i,M}}{n_M} - \frac{tf_{i,F}}{n_F} \quad (9.1)$$

где $tf_{i,M}$ - абсолютная частота i -го слова в категории M ; n_M - длина (в словах) всех сообщений, принадлежащих классу M (аналогично для противоположного класса F).

c. Определение тем, обсуждаемых в тексте (LDA, PLSA)

3. Также можно применить критерии выбора признаков, играющие роль фильтра.

Для каждого термина они вычисляет оценку распределения по двум категориям.

В качестве критериев можно использовать: отношение шансов (OR), хи-квадрат и коэффициент корреляции. Каждая из этих функций находит, как правило, различные термины, характерные для данной категории.

4. Использование методов глубокого обучения(CNN, LSTM, Bi-LSTM, GRU) и готовых моделей (GloVe, BERT и т.п.)

Отчет о проделанной работе

Подготовка данных

Сначала планировалось использовать данные из задания на [PAN at CLEF 2017](#), PAN at FIRE Track on Cross-genre Gender Identification in Russian.

Однако оказалось, что ссылки на датасеты не работают, а запрос к организаторам о предоставлении возможности получить доступ к данным, остался без ответа.

К счастью, удалось найти источник - общую базу данных корпусов, из которых формировались датасеты для задания на конференции (<https://rusidiolect.rusprofilinglab.ru/>).

Поскольку структура датасетов, как тренировочных, так и тестовых, подробно описана в [d8], в итоге, для модели и ее оценки, получилось подготовить датасеты, аналогичные данным из задания на RusProfiling PAN '17.

Набор данных Twitter:

Всего

- Использовались сообщения 1000 пользователей, по 500 каждого гендера
- Все сообщения от одного пользователя были объединены и считались одним текстом
- Длина текста одного пользователя - в среднем, 500 слов или, в среднем, 1500 слов
- Весь датасет был разделен на два: *набор данных для обучения* и *набор данных для тестирования*

Набор данных для обучения модели

- Содержал тексты 600 пользователей, по 300 каждого гендера

Набор данных для тестирования одножанровых текстов

- Содержал тексты 400 пользователей, по 200 каждого гендера

Набор данных Facebook:

Набор данных для тестирования одножанровых текстов

- Содержал тексты 228 пользователей, по 114 каждого гендера
- Все сообщения от одного пользователя были объединены и считались одним текстом
- Размер текста одного пользователя - около 1000 слов

Поскольку источником для упомянутых выше датасетов *Twitter* и *Facebook* была одна база текстов, в которой не был указан источник, деление на разные соцсети проводилось по единственному упомянутому в [d8] признаку - размеру текста одного пользователя. Поэтому допустимо считать все тексты из базы одножанровыми.

Набор данных эссе (Essays):

Набор данных для тестирования разножанровых текстов

- Содержал тексты 370 пользователей, по 185 каждого гендера
- Каждый пользователь - автор одного или двух текстов
- Темами текстов были письмо другу/подруге, описание фотографии, письмо сотруднице, пытающейся убедить ее нанять респондента
- Длина текста одного пользователя - в среднем, 150 слов

Набор данных обзоров (Reviews):

Набор данных для тестирования разножанровых текстов

- Содержал тексты 776 пользователей, по 388 каждого гендера
- Каждый пользователь - автор одного текста
- Длина текста одного пользователя - в среднем, 80 слов

Набор данных, имитирующий гендер (Gender imitation):

Набор данных для тестирования разножанровых текстов

- Содержал тексты 94 пользователей, по 47 каждого гендера
- Каждый пользователь - автор трех текстов
- Тема у всех трех текстов одна (из набора тем)
 - первый текст написан обычным способом для того, кто его писал (без всякого обмана)
 - второй написан от лица противоположного пола («имитация»)
 - третий написан так, чтобы личный стиль письма не был распознан, как если бы автором являлось другое лицо того же пола («запутывание»)
- Все тексты от одного пользователя были объединены и считались одним текстом
- Длина текста одного пользователя - от 80 до 150 слов

Для экспериментов с моделями, обученных на разных по объему наборах данных, были собраны два дополнительных одножанровых датасета.

В Среднем наборе данных для обучения использовался весь Набор данных *Twitter* (*training+test*).

В Большом наборе данных для обучения использовался объединенный Набор данных *Twitter(training+test)* с Набором данных *Facebook*.

Средний набор данных для обучения:

- Использовались сообщения 1000 пользователей, по 500 каждого гендера
- Все сообщения от одного пользователя были объединены и считались одним текстом
- Длина текста одного пользователя - в среднем, от 500 слов или, в среднем, 1500 слов

Большой набор данных для обучения:

- Использовались сообщения 1228 пользователя, по 614 каждого гендера
- Все сообщения от одного пользователя были объединены и считались одним текстом
- Длина текста одного пользователя - равномерно, от, в среднем, 500 слов до, в среднем, 1500 слов

Функция потерь

Поскольку в целевом объекте присутствуют два класса, в качестве функции потерь используется *Binary Cross Entropy Loss*:

$$BCE = -\frac{1}{N} \sum_{i=0}^N (y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i))$$

Построение модели

В модели классификации текста - три полносвязных слоя.

В качестве базового BertLayer использовалась *Русская модель SBERT NLU*.

Русская модель SBERT NLU – это *BERT-large NLU*, состоящая из 24 слоев и содержащая 426.9 миллионов параметров. (Подробнее про создание Русской модели SBERT NLU - <https://www.youtube.com/watch?v=W3ko2fWWSQI>).

Для tensorflow русская модель *SBERT NLU* доступна для скачивания по ссылке https://storage.googleapis.com/bert_resources/sbert_nlu/sbert_cased_L-24_H-1024_A-16_tf.zip

С последнего слоя *BERT* берутся усреднения эмбеддингов токенов (используется *masked mean pooling*).

Далее идет скрытый слой (Dense) с 1024 нейронами и активационной функцией *relu*.

Для регуляризации, между последним скрытым слоем и выходным слоем, используется *Dropout = 0.25*.

Последний слой (классификатор) состоит из полносвязного слоя с двумя выходными нейронами и функцией активации *sigmoid*.

Оптимизационный алгоритм - *Adam*, значение *learning_rate = 10⁻⁵*.

Качество предсказаний модели будем оценивать по двум метрикам - *accuracy* и *F1*-мере.

Обучение

Параметры обучения:

- Базовый BertLayer — русская модель SBERT NLU;
- batch size = 24;
- seq_len = 256;
- N_tune_lrs = 24;
- Optimizer — Adam;
- Learning rate = 1e-5;
- Метрики — F1, Accurasy;
- Количество эпох = 3.

Model summary представлена ниже:

Model:

| Layer (type) | Output Shape | Param # |
|-----------------------------------|--------------|-----------|
| input_1 (InputLayer) | [(None, 1)] | 0 |
| bert_layer (BertLayer) | (None, 1024) | 426908672 |
| dense (Dense) | (None, 1024) | 1049600 |
| dropout (Dropout) | (None, 1024) | 0 |
| dense_1 (Dense) | (None, 2) | 2050 |
| Total params: 427,960,322 | | |
| Trainable params: 303,361,026 | | |
| Non-trainable params: 124,599,296 | | |

Модели просчитывались на iMac (4,2 GHz Intel Core i7, 64 ГБ 2400 MHz DDR4) без GPU.

Эксперимент

Для проведения эксперимента были подготовлены три модели с одинаковыми параметрами, но обученные на разных по размеру датасетах:

- **RuSb_base** - на 600 сообщений пользователей *Twitter*, поровну от каждого гендера

- **RuSb_mid** - на 1000 сообщений пользователей *Twitter*, поровну от каждого гендера (Средний набор данных для обучения)
- **RuSb_big** - на 1228 сообщений пользователей *Twitter*, поровну от каждого гендера (Большой набор данных для обучения)

Еще три модели, основанные на описанных выше, учитывали *стилометрические признаки* гендера *S* для русского языка.

Стилометрические признаки гендера *S* для русского языка:

В русском языке формы единственного числа глаголов прошедшего времени меняются по родам (мужские формы единственного числа имеют окончание “- л”, а индикатором женских форм единственного числа является окончание “- ла”).

Также, различаются по родам возвратные глаголы прошедшего времени - для женского рода имеют окончание “-лась”, а для мужского - “-лся”.

Описанные особенности глаголов использовались в сочетании с подлежащим предложения, если подлежащее было местоимением первого лица единственного числа “я” (или “Я”), и если подлежащее предстояло от глагола не более, чем в пределах окна из 4 слов.

Таким образом, всего было использовано четыре признака: два для мужского рода - “Я(я) [0-4] -л” и “Я(я) [0-4] -лся”, и два для женского рода - “Я(я) [0-4] -ла” и “Я(я) [0-4] -лась”.

Результаты и наблюдения

В таблице указаны полученные результаты (*accuracy*) разных моделей, в том числе с учетом стилометрических признаков для тестовых одно- и многожанровых наборов. Напомню, результат для *Facebook*, исходя из построения датасетов, можно считать одножанровым.

| | Twitter | Facebook | Essays | Reviews | Gender imitation |
|-----------------------------|---------|----------|--------|---------|------------------|
| RuSb_base | 0.87 | 0.85 | 0.86 | 0.80 | 0.93 |
| RuSb_base+stylometry | 0.90 | 0.89 | 0.87 | 0.80 | |
| RuSb_mid | | 0.85 | | | |
| RuSb_mid+stylometry | | 0.89 | | | |
| RuSb_big | | | 0.86 | 0.83 | 0.95 |
| RuSb_big+stylometry | | | 0.86 | 0.83 | |

1. Стилометрические признаки для *Reviews* не повлияли на результат, в то время, как в остальных случаях, как одно-, так и многожанровых, есть небольшое увеличение.

Возможно, это обусловлено краткостью сообщений в *Reviews* (до 80 слов), где рассматриваемые стилометрическими признаками формы изложения употребляются редко.

2. Самые низкие результаты в *Reviews* - также, на мой взгляд, по причине краткости сообщений.
3. Увеличение обучающего датасета не сработало в двух случаях: *Facebook* и *Essays*. Если в первом, одножанровом случае (*Facebook*), можно предположить, что размер базового датасета оказался достаточен, то в случае с *Essays*, логичным выглядит предположение, что причиной явилось ограниченность тем, заданных в качестве основы для текстов пользователей.

Более подробная информация по метрикам всех экспериментов приведена ниже:

One-genre Gender Detection (Russian)

Twitter

| RuSb_base | | | | | |
|-----------|--------------|-----------|--------|----------|---------|
| | | precision | recall | f1-score | support |
| | 0 | 0.88 | 0.85 | 0.87 | 200 |
| | 1 | 0.86 | 0.88 | 0.87 | 200 |
| | accuracy | | | 0.87 | 400 |
| | macro avg | 0.87 | 0.87 | 0.87 | 400 |
| | weighted avg | 0.87 | 0.87 | 0.87 | 400 |

| RuSb_base+stylometry | | | | | |
|----------------------|--------------|-----------|--------|----------|---------|
| | | precision | recall | f1-score | support |
| | 0 | 0.89 | 0.92 | 0.90 | 200 |
| | 1 | 0.91 | 0.89 | 0.90 | 200 |
| | accuracy | | | 0.90 | 400 |
| | macro avg | 0.90 | 0.90 | 0.90 | 400 |
| | weighted avg | 0.90 | 0.90 | 0.90 | 400 |

Facebook

| RuSb_base | | | | | |
|-----------|--|-----------|--------|----------|---------|
| | | precision | recall | f1-score | support |

| | | | | |
|--------------|------|------|------|-----|
| 0 | 0.86 | 0.85 | 0.85 | 110 |
| 1 | 0.85 | 0.86 | 0.86 | 110 |
| accuracy | | | 0.85 | 220 |
| macro avg | 0.85 | 0.85 | 0.85 | 220 |
| weighted avg | 0.85 | 0.85 | 0.85 | 220 |

RuSb_base+stylometry

| | | | | |
|--------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.88 | 0.90 | 0.89 | 110 |
| 1 | 0.90 | 0.88 | 0.89 | 110 |
| accuracy | | | 0.89 | 220 |
| macro avg | 0.89 | 0.89 | 0.89 | 220 |
| weighted avg | 0.89 | 0.89 | 0.89 | 220 |

RuSb_mid

| | | | | |
|--------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.86 | 0.84 | 0.85 | 110 |
| 1 | 0.84 | 0.86 | 0.85 | 110 |
| accuracy | | | 0.85 | 220 |
| macro avg | 0.85 | 0.85 | 0.85 | 220 |
| weighted avg | 0.85 | 0.85 | 0.85 | 220 |

RuSb_mid+stylometry

| | | | | |
|---|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.89 | 0.89 | 0.89 | 110 |
| 1 | 0.89 | 0.89 | 0.89 | 110 |

| | | | | |
|--------------|------|------|------|-----|
| accuracy | | | 0.89 | 220 |
| macro avg | 0.89 | 0.89 | 0.89 | 220 |
| weighted avg | 0.89 | 0.89 | 0.89 | 220 |

Cross-genre Gender Detection (Russian)

Essays

RuSb_base

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.85 | 0.88 | 0.86 | 185 |
| 1 | 0.88 | 0.84 | 0.86 | 185 |
| accuracy | | | 0.86 | 370 |
| macro avg | 0.86 | 0.86 | 0.86 | 370 |
| weighted avg | 0.86 | 0.86 | 0.86 | 370 |

RuSb_base+stylometry

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.86 | 0.88 | 0.87 | 185 |
| 1 | 0.88 | 0.86 | 0.87 | 185 |
| accuracy | | | 0.87 | 370 |
| macro avg | 0.87 | 0.87 | 0.87 | 370 |
| weighted avg | 0.87 | 0.87 | 0.87 | 370 |

RuSb_big

| | precision | recall | f1-score | support |
|---|-----------|--------|----------|---------|
| 0 | 0.86 | 0.85 | 0.86 | 185 |
| 1 | 0.85 | 0.86 | 0.86 | 185 |

| | | | | |
|--------------|------|------|------|-----|
| accuracy | | | 0.86 | 370 |
| macro avg | 0.86 | 0.86 | 0.86 | 370 |
| weighted avg | 0.86 | 0.86 | 0.86 | 370 |

RuSb_big+stylometry

| | | | | |
|--------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.87 | 0.86 | 0.86 | 185 |
| 1 | 0.86 | 0.87 | 0.87 | 185 |
| accuracy | | | 0.86 | 370 |
| macro avg | 0.86 | 0.86 | 0.86 | 370 |
| weighted avg | 0.86 | 0.86 | 0.86 | 370 |

Reviews

RuSb_base

| | | | | |
|--------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.95 | 0.64 | 0.77 | 388 |
| 1 | 0.73 | 0.97 | 0.83 | 388 |
| accuracy | | | 0.80 | 776 |
| macro avg | 0.84 | 0.80 | 0.80 | 776 |
| weighted avg | 0.84 | 0.80 | 0.80 | 776 |

RuSb_base+stylometry

| | | | | |
|-----------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.95 | 0.64 | 0.77 | 388 |
| 1 | 0.73 | 0.97 | 0.83 | 388 |
| accuracy | | | 0.80 | 776 |
| macro avg | 0.84 | 0.80 | 0.80 | 776 |

| | | | | |
|--------------|------|------|------|-----|
| weighted avg | 0.84 | 0.80 | 0.80 | 776 |
|--------------|------|------|------|-----|

RuSb_big

| | | | | |
|--------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.95 | 0.70 | 0.81 | 388 |
| 1 | 0.76 | 0.97 | 0.85 | 388 |
| accuracy | | | 0.83 | 776 |
| macro avg | 0.86 | 0.83 | 0.83 | 776 |
| weighted avg | 0.86 | 0.83 | 0.83 | 776 |

RuSb_big+stylometry

| | | | | |
|--------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.95 | 0.70 | 0.81 | 388 |
| 1 | 0.76 | 0.97 | 0.85 | 388 |
| accuracy | | | 0.83 | 776 |
| macro avg | 0.86 | 0.83 | 0.83 | 776 |
| weighted avg | 0.86 | 0.83 | 0.83 | 776 |

Gender imitation

RuSb_base

| | | | | |
|--------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.92 | 0.94 | 0.93 | 47 |
| 1 | 0.93 | 0.91 | 0.92 | 47 |
| accuracy | | | 0.93 | 94 |
| macro avg | 0.93 | 0.93 | 0.93 | 94 |
| weighted avg | 0.93 | 0.93 | 0.93 | 94 |

RuSb_big

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.96 | 0.94 | 0.95 | 47 |
| 1 | 0.94 | 0.96 | 0.95 | 47 |
| accuracy | | | 0.95 | 94 |
| macro avg | 0.95 | 0.95 | 0.95 | 94 |
| weighted avg | 0.95 | 0.95 | 0.95 | 94 |

4. Обращает внимание дисбаланс *precision* и *recall* во всех моделях на датасете *Reviews* - высокая точность отличать класс 0 (женский) при невысоком проценте обнаруживать сам класс, и наоборот для класса 1 (мужской).

Сравнение с предыдущими результатами

В таблицах ниже приведен рейтинг моделей *RuSb_base*, *RuSb_base+stylometry*, *RuSb_big*, *RuSb_big+stylometry* в сравнении с другими моделями, рассмотренными в Обзоре. Сравнение по *Facebook* не проводилось, так как не удалось найти релевантный тестовый датасет для этой категории.

TABLE V. One-genre Gender Detection (Russian)

| Work | Publishing Year | Corpus | Features | Method used | Result |
|-----------------------------|-----------------|---|----------------------|-------------------|-------------|
| RuSb_base+stylometry | | PAN FIRE '17 [d8] | Stylometric features | Russian SBERT | 0.90 |
| RuSb_base | | PAN FIRE '17 [d8] | | Russian SBERT | 0.87 |
| Korshunov [r6] | 2013 | self-made | word (3-grams) | SVM | 0.86 |
| Sboev et al [r4] | 2019 | RusProfilihg [d11] PAN FIRE '17 [d8] | char n-grams | Gradient Boosting | 0.79 |
| Сбоев и др. [r1] | 2020 | LiveJournal [d10] | | GRU, CVAE | 0.76 |
| Markov et al - CIC3 [r2] | 2017 | PAN FIRE '17 [d8] | statistical | | 0.6825 |

| | | | | | |
|--------------------------|------|-------------------|--|---------------|--------|
| LDR [d12] | 2017 | PAN FIRE '17 [d8] | probability distribution of occurrence of tdoc's words in the different classes. | | 0.6759 |
| Markov et al - CIC2 [r2] | 2017 | PAN FIRE '17 [d8] | BOW, word (suffix 3-grams), tf-idf | SVM | 0.6650 |
| Bhargava et al [r3] | 2017 | PAN FIRE '17 [d8] | POS, rule-based classification | LSTM, Bi-LSTM | 0.6525 |
| Markov et al - CIC1 [r2] | 2017 | PAN FIRE '17 [d8] | POS combination, tf-idf | SVM | 0.6525 |

TABLE VI. Cross-genre Gender Detection (Russian)

| Work | Publishing Year | Corpus | Features | Method used | Test Corpus | Result |
|-----------------------------|-----------------|---|----------------------|---------------|-------------|-------------|
| RuSb_base+stylometry | | PAN FIRE '17 [d8] (Twitter - training) | Stylometric features | Russian SBERT | Essays | 0.87 |
| RuSb_base | | PAN FIRE '17 [d8] (Twitter - training) | | Russian SBERT | Essays | 0.86 |
| RuSb_big | | RusProfiling [d11] PAN FIRE '17 [d8] | | Russian SBERT | Essays | 0.86 |
| RuSb_big+stylometry | | RusProfiling [d11] PAN FIRE '17 [d8] | Stylometric features | Russian SBERT | Essays | 0.86 |
| LDR [d12] | 2017 | PAN FIRE | stylometric | | Essays | 0.8141 |

| | | | | | | |
|-----------------------------|------|---|---|----------------------|---------|-------------|
| | | '17 [d8] (Twitter - training) | analysis | | | |
| Bhargava et al [r3] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | POS, rule- based classification | LSTM, Bi-LSTM | Essays | 0.7838 |
| Vinayan et al [r5] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | exotic stat (average word length, URL usage, etc), tf-idf | SVM | Essays | 0.6811 |
| RuSb_big+stylometry | | RusProfiling [d11] PAN FIRE '17 [d8] | Stylometric features | Russian SBERT | Reviews | 0.83 |
| RuSb_big | | RusProfiling [d11] PAN FIRE '17 [d8] | | Russian SBERT | Reviews | 0.83 |
| RuSb_base+stylometry | | PAN FIRE '17 [d8] (Twitter - training) | Stylometric features | Russian SBERT | Reviews | 0.80 |
| RuSb_base | | PAN FIRE '17 [d8] (Twitter - training) | | Russian SBERT | Reviews | 0.80 |
| Sboev et al [r4] | 2019 | PAN FIRE '17 [d8] (Twitter - training) | char n- grams | Gradient Boosting | Reviews | 0.79 |
| LDR [d12] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | stylometric analysis | | Reviews | 0.72 |
| Markov et al - CIC3 [r2] | 2017 | PAN FIRE '17 [d8] | statistical | | Reviews | 0.6186 |

| | | | | | | |
|-----------------------------|------|--|--------------------------------|---------------|------------------|-------------|
| | | (Twitter - training) | | | | |
| Markov et al - CIC1 [r2] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | POS combination, tf-idf | SVM | Reviews | 0.5979 |
| Bhargava et al [r3] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | POS, rule-based classification | LSTM, Bi-LSTM | Reviews | 0.5786 |
| RuSb_base+stylometry | | PAN FIRE '17 [d8] (Twitter - training) | Stylometric features | Russian SBERT | Gender imitation | 0.95 |
| RuSb_base | | PAN FIRE '17 [d8] (Twitter - training) | | Russian SBERT | Gender imitation | 0.93 |
| Bhargava et al [r3] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | POS, rule-based classification | LSTM, Bi-LSTM | Gender imitation | 0.6596 |
| LDR [d12] | 2017 | PAN FIRE '17 [d8] (Twitter - training) | stylometric analysis | | Gender imitation | 0.6383 |

Выводы

1. Использование, в качестве базовой, *Русской модели SBERT NLU* с дальнейшим fine-tuning под задачу классификации гендера, дает модель, превышающую по результатам модели, рассмотренные выше в Обзоре - как с традиционным подходом, так и модели, использующие методы глубокого обучения.
2. Результаты предсказания в гендерном профилировании на одножанровых и многожанровых датасетах не имеют существенных отличий, за исключением специфичных требований к текстам в случае отличия жанра тестируемой выборки от того, на котором происходило обучение (размер сообщений, ограниченность тематики и т.п.).

3. Увеличение обучающего датасета дает незначительное увеличение точности предсказания на одножанровых и некоторых многожанровых датасетах. При наличии требования об ограничении размера сообщений в датасете небольшим значением (менее 100 слов), в случае отличия жанра тестируемой выборки от того, на котором происходило обучение, точность не меняется.
4. Использование стилометрических признаков позволяет улучшить результаты предсказания модели, построенной на *Русской модели SBERT NLU*, в гендерном профилировании на одножанровых и многожанровых датасетах. Однако, прирост в результатах незначителен.

Notebooks

Тексты программ на github - https://github.com/dbadeev/gender_profiling

-
-
1. [t1] Francisco Rangel, Paolo Rosso, Ben Verhoeven, Walter Daelemans, Martin Potthast, and Benno Stein. Overview of the 4th Author Profiling Task at PAN 2016: Cross-Genre Evaluations. In Krisztian Balog, Linda Cappellato, Nicola Ferro, and Craig Macdonald, editors, CLEF 2016 Evaluation Labs and Workshop – Working Notes Papers, 5-8 September, Évora, Portugal, September 2016. CEUR-WS.org
 2. [t2] Francisco Rangel, Paolo Rosso, Martin Potthast, and Benno Stein. [Overview of the 5th Author Profiling Task at PAN 2017: Gender and Language Variety Identification in Twitter](#). In Linda Cappellato, Nicola Ferro, Lorraine Goeuriot, and Thomas Mandl, editors, CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, 11-14 September, Dublin, Ireland, September 2017. CEUR-WS.org
 3. [t3] [Francisco Rangel, Paolo Rosso. Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling in Twitter](#)
 4. [m1] J.F. Burrows, Not unless you ask nicely: the interpretative Nexus between analysis and information. Lit. Linguis. Comput. 7(1), 91–109 (1992)
 5. [m2] J.F. Burrows, Delta: a measure of stylistic difference and a guide to likely authorship. Lit. Linguis. Comput. 17(3), 267–287 (2002)
 6. [m3] J. Savoy, Comparative evaluation of term selection functions for authorship attribution. Digit. Scholarsh. Humanit. 30(2), 246–261 (2015)
 7. [d1] Linda Cappellato, Nicola Ferro, Martin Halvey, and Wessel Kraaij, editors. CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/> Vol-1180/, 2014.
 8. [d2] Ben Verhoeven and Walter Daelemans. Clips stylometry investigation (csi) corpus: a dutch corpus for the detection of age, gender, personality, sentiment and deception in text. In 9th International Conference on Language Resources and Evaluation (LREC 2014), 2014.
 9. [d3] Ben Verhoeven, Walter Daelemans, and B. Plank. Twisty: a multilingual twitter stylometry corpus for gender and personality profiling. In 10th International Conference on Language Resources and Evaluation (LREC 2016), 2016.

10. [d4] Zaghouani W, Charf A. Arapâ tweet: a large multiâ dialect twitter corpus for gender, age and language variety identifcation. In: Proceedings of the 11th international conference on language resources and evaluation (LREC), Miyazaki, Japan; 2018.
11. [d5] B. Tayfun Kucukyilmaz, B. Barla Cambazoglu, Cevdet Aykanat, and Fazli Can. "Chat mining for gender prediction." In International conference on advances in information systems, pp. 274-283. Springer, Berlin, Heidelberg, 2006
12. [d6] Francisco Rangel, Paolo Rosso, Martin Potthast, and Benno Stein. [Overview of the 5th Author Profiling Task at PAN 2017: Gender and Language Variety Identification in Twitter](#). In Linda Cappellato, Nicola Ferro, Lorraine Goeuriot, and Thomas Mandl, editors, *CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, 11-14 September, Dublin, Ireland*, September 2017. CEUR-WS.org
13. [d7] [Francisco Rangel, Paolo Rosso. Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling in Twitter](#)
14. [d8] Litvinova T, Rangel F, Rosso P, Seredin P, Litvinova O. Overview of the rusprofiling pan at fre track on cross-genre gender identifcation in Russian. In: Notebook papers of FIRE 2017, FIRE-2017, Bangalore, India, December 8–11, CEUR Workshop Proceedings. CEUR-WS.org, vol 2036; 2017. pp 1–7.
15. [d9] T. Litvinova, O. Litvinlova, O. Zagorovskaya, P. Seredin, A. Sboev, and O. Romanchenko, ""ruspersonality": A russian corpus for authorship profiling and deception detection," in Intelligence, Social Media and Web (ISMW FRUCT), 2016 International FRUCT Conference on (IEEE, 2016), pp. 1–7.
16. [d10] *Litvinova T.A., Sboev A.G., Panicheva P.V.* Profiling the Age of Russian Bloggers // Conference on Artificial In- telligence and Natural Language. Springer, Cham, 2018. C. 167–177.
17. [d11] T. Litvinova, O. Litvinlova, O. Zagorovskaya, P. Seredin, A. Sboev, and O. Romanchenko, ""ruspersonality": A russian corpus for authorship profiling and deception detection," in Intelligence, Social Media and Web (ISMW FRUCT), 2016 International FRUCT Conference on (IEEE, 2016), pp. 1–7.
18. [d12] F. Rangel, P. Rosso, and M. Franco-Salvador. A low dimensionality representation for language variety identification. In 17th International Conference on Intelligent Text Processing and Computational Linguistics, CICLing. Springer-Verlag, LNCS, arXiv:1705.10754, 2016.
19. [d13] M. Fatima, K. Hasan, S. Anwar and R.M.A. Nawab, 799 Multilingual author profiling on facebook, *Information Pro- cessing & Management* 53(4) (2017), 886–904.
20. [d14] M. Fatima, S. Anwar, A. Naveed, W. Arshad, R.M.A. 795 Nawab, M. Iqbal and A. Masood, Multilingual sms-based 796 author profiling: Data and methods, *Natural Language 797 Engineering* 24 (5) (2018), 695-724.
21. [x1] Matej Martinc, Iza Škrjanec, Katja Zupan, and Senja Pollak. [PAN 2017: Author Profiling - Gender and Language Variety Prediction—Notebook for PAN at CLEF 2017](#). In Linda Cappellato, Nicola Ferro, Lorraine Goeuriot, and Thomas Mandl, editors, *CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, 11-14 September, Dublin, Ireland*, September 2017. CEUR-WS.org.

22. [x2] Yasuhide Miura, Tomoki Taniguchi, Motoki Taniguchi, and Tomoko Ohkuma. [Author Profiling with Word+Character Neural Attention Network—Notebook for PAN at CLEF 2017](#). In Linda Cappellato, Nicola Ferro, Lorraine Goeuriot, and Thomas Mandl, editors, *CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, 11-14 September, Dublin, Ireland*, September 2017. CEUR-WS.org.
23. [x3] Angelo Basile, Gareth Dwyer, Maria Medvedeva, Josine Rawee, Hessel Haagsma, and Malvina Nissim. [N-GrAM: New Groningen Author-profiling Model—Notebook for PAN at CLEF 2017](#). In Linda Cappellato, Nicola Ferro, Lorraine Goeuriot, and Thomas Mandl, editors, *CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, 11-14 September, Dublin, Ireland*, September 2017. CEUR-WS.org.
24. [x4] Juan Pizarro. [Using N-grams to detect Bots on Twitter](#). In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, *CLEF 2019 Labs and Workshops, Notebook Papers*, September 2019. CEUR-WS.org
25. [x5] Mahendrakar Srinivasarao and Siddharth Manu. [Bots and Gender Profiling using Character and Word N-Grams](#). In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, *CLEF 2019 Labs and Workshops, Notebook Papers*, September 2019. CEUR-WS.org
26. [x6] Youngjun Joo and Inchon Hwang. [Steve Martin at SemEval-2019 Task 4: Ensemble Learning Model for Detecting Hyperpartisan News](#). In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 990–994, June 2019. Association for Computational Linguistics.
27. [x7] Régis Goubin, Dorian Lefeuvre, Alaa Alhamzeh, Jelena Mitrović, Előd Egyed-Zsigmond, and Leopold Ghemmogne Fossi. [Bots and Gender Profiling using a Multi-layer Architecture](#). In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, *CLEF 2019 Labs and Workshops, Notebook Papers*, September 2019. CEUR-WS.org
28. [x8] Alex I. Valencia Valencia, Helena Gomez Adorno, Christopher Stephens Rhodes, and Gibran Fuentes Pineda. [Bots and Gender Identification Based on Stylometry of Tweet Minimal Structure and n-grams Model](#). In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, *CLEF 2019 Labs and Workshops, Notebook Papers*, September 2019. CEUR-WS.org.
29. [x9] Marco Polignano, Marco Giuseppe de Pinto, Pasquale Lops, and Giovanni Semeraro. [Identification Of Bot Accounts In Twitter Using 2D CNNs On User-generated Contents](#). In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, *CLEF 2019 Labs and Workshops, Notebook Papers*, September 2019. CEUR-WS.org.
30. [x10] M.A. Ashraf, R.M.A Nawab, and F. Nie, Author profiling on bi-lingual tweets. *Journal of Intelligent & Fuzzy Systems*, (Preprint), pp.1-11.2019.
31. [x11] Flurin Gishamer. [Using Hashtags and POS-Tags for Author Profiling](#). In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, *CLEF 2019 Labs and Workshops, Notebook Papers*, September 2019. CEUR-WS.org.
32. [x12] B. Tayfun Kucukyilmaz, B. Barla Cambazoglu, Cevdet Aykanat, and Fazli Can. "Chat mining for gender prediction." In *International conference on advances in information systems*, pp. 274-283. Springer, Berlin, Heidelberg, 2006

33. [r1] Сбоев А.Г., Рыбка Р.Б., Грязнов А.В., Молошников И.А. ГЕНЕРАТИВНО-ДИСКРИМИНАТИВНАЯ НЕЙРОСЕТЕВАЯ МОДЕЛЬ ДЛЯ ЗАДАЧИ АВТОРСКОГО ПРОФИЛИРОВАНИЯ, Вестник Национального исследовательского ядерного университета МИФИ. 2020. Т. 9. № 1. С. 50-57
34. [r2] I. Markov, H. Gomez-Adorno, G. Sidorov, and A. Gelbukh. The winning approach to cross-genre gender identification in russian at rusprofiling 2017. In Working Notes for PAN-RUSProfiling at FIRE'17. Workshops Proceedings of the 9th International Forum for Information Retrieval Evaluation (Fire'17), Bangalore, India. CEUR-WS.org, 2017
35. [r3] R. Bhargava, G. Goel, A. Shah, and Y. Sharma. Gender identification in russian texts. In Working Notes for PAN-RUSProfiling at FIRE'17. Workshops Proceedings of the 9th International Forum for Information Retrieval Evaluation (Fire'17), Bangalore, India. CEUR-WS.org, 2017.
36. [r4] Sboev, D. Gudovskikh, . Moloshnikov, R. Rybka. A gender identification of text author in mixture of Russian multi-genre texts with distortions on base of data-driven approach using machine learning models. AIP Conference Proceedings 2116, 270006 (2019)
37. [r5] V. Vinayan, N. J.R., H. NB, A. Kumar M, and S. K P. Amritanlp@pan-rusprofiling: Author profiling using machine learning techniques. In Working Notes for PAN-RUSProfiling at FIRE'17. Workshops Proceedings of the 9th International Forum for Information Retrieval Evaluation (Fire'17), Bangalore, India. CEUR-WS.org, 2017
38. [r6] Korshunov A., B. I. (2013). Detection of demographic attributes of microblog users. Proceedings of the Institute for System Programming Volume 25.
39. [s1] Mart Busger op Vollenbroek, Talvany Carlotto, Tim Kreutz, Maria Medvedeva, Chris Pool, Johannes Bjerva, Hessel Haagsma, and Malvina Nissim. [GronUP: Groningen User Profiling—Notebook for PAN at CLEF 2016](#). In Krisztian Balog, Linda Cappellato, Nicola Ferro, and Craig Macdonald, editors, *CLEF 2016 Evaluation Labs and Workshop – Working Notes Papers, 5-8 September, Évora, Portugal, September 2016*. CEUR-WS.org.
40. [s2] Madhulika Agrawal and Teresa Gonçalves. [Age and Gender Identification using Stacking for Classification—Notebook for PAN at CLEF 2016](#). In Krisztian Balog, Linda Cappellato, Nicola Ferro, and Craig Macdonald, editors, *CLEF 2016 Evaluation Labs and Workshop – Working Notes Papers, 5-8 September, Évora, Portugal, September 2016*. CEUR-WS.org
41. [s3] Pashutan Modaresi, Matthias Liebeck, and Stefan Conrad. [Exploring the Effects of Cross-Genre Machine Learning for Author Profiling in PAN 2016—Notebook for PAN at CLEF 2016](#). In Krisztian Balog, Linda Cappellato, Nicola Ferro, and Craig Macdonald, editors, *CLEF 2016 Evaluation Labs and Workshop – Working Notes Papers, 5-8 September, Évora, Portugal, September 2016*. CEUR-WS.org.
42. [s4] Ivan Bilan and Desislava Zhekova. [CAPS: A Cross-genre Author Profiling System—Notebook for PAN at CLEF 2016](#). In Krisztian Balog, Linda Cappellato, Nicola Ferro, and Craig Macdonald, editors, *CLEF 2016 Evaluation Labs and Workshop – Working Notes Papers, 5-8 September, Évora, Portugal, September 2016*. CEUR-WS.org.
43. [s5] Roy Bayot and Teresa Gonçalves. [Author Profiling using SVMs and Word Embedding Averages—Notebook for PAN at CLEF 2016](#). In Krisztian Balog, Linda Cappellato, Nicola

Ferro, and Craig Macdonald, editors, *CLEF 2016 Evaluation Labs and Workshop – Working Notes Papers*, 5-8 September, Évora, Portugal, September 2016. CEUR-WS.org.