Style Change Detection

Зуева Надежда

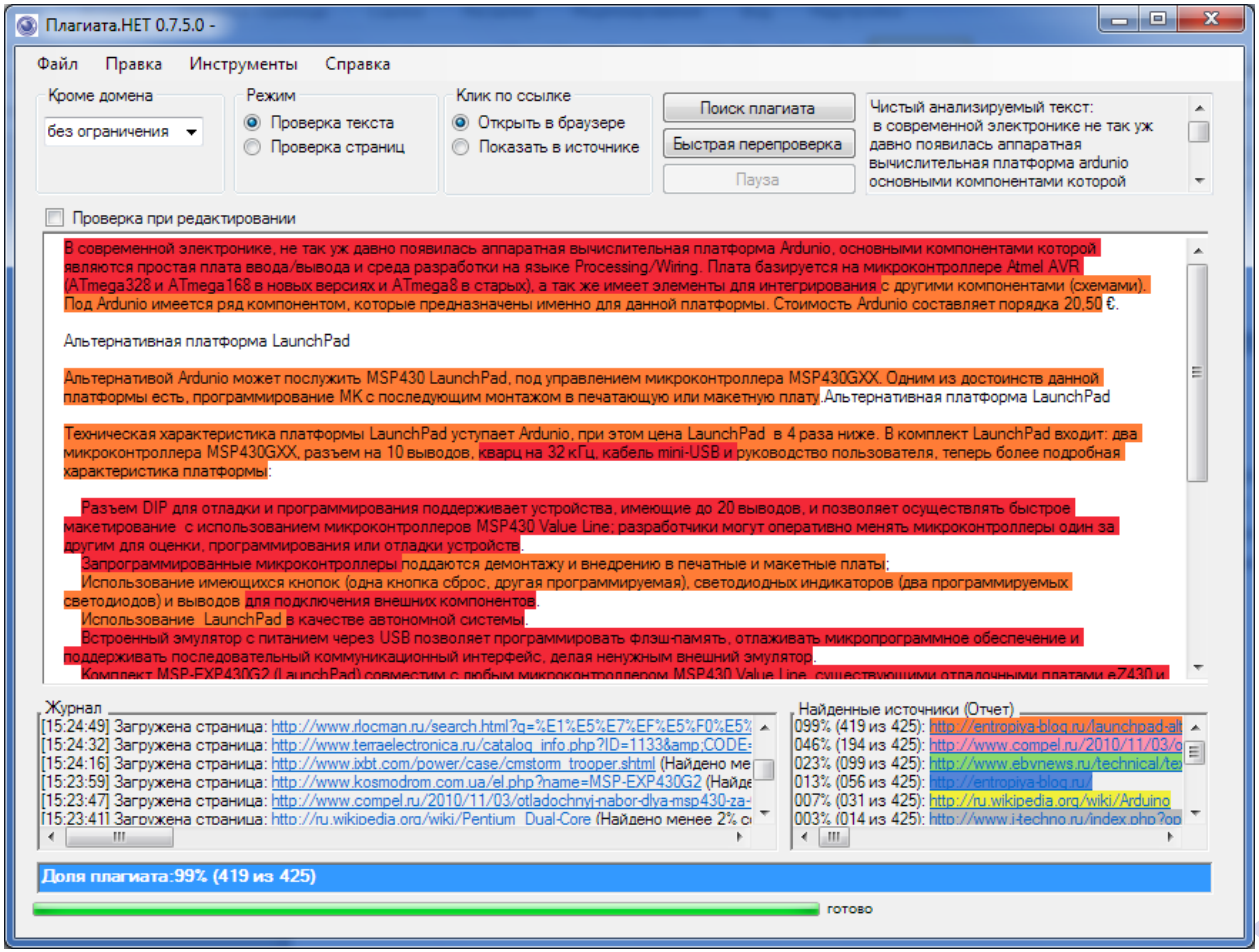
March 2018

Введение. Цели исследования

1. Написание статей
2. Достоверность исторических документов
3. Составление профиля автора и восстановление его портрета

Существующие методы

1. Построение вектора признаков и поиск отклонений выше некоторой нормы
2. Ранние методы, не содержащие машинное обучение, реагирующие на т.н. **стоп-слова**



Литература

1. Stein, B., Barrón Cedeño, L.A., Eiselt, A., Potthast, M., Rosso, P.: Overview of the 3rd international competition on plagiarism detection. In: CEUR Workshop Proceedings. CEUR Workshop Proceedings (2011)
2. http://pan.webis.de/clef18/pan18-web/author-identification.html
3. https://pdfs.semanticscholar.org/1011/6d82a8438c78877a8a142be47c4ee86
4. https://arxiv.org/pdf/1701.06547.pdf
5. Zechner, M., Muhr, M., Kern, R., Granitzer, M.: External and intrinsic plagiarism detection using vector space models. Proc. SEPLN. vol. 32 (2009)

Постановка задачи

Если встречен подозрительный на плагиат документ, необходимо определить, написан ли он одним автором или содержит нелегитимные заимствования Два подхода:

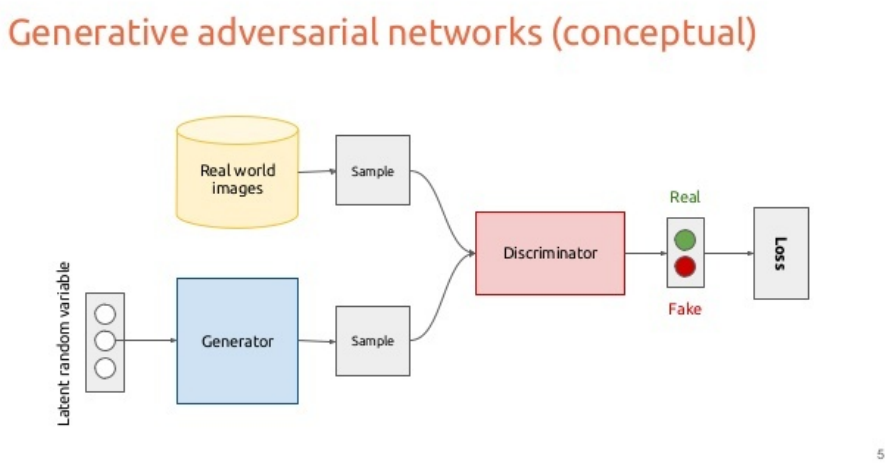
1. Внешний. Для поиска внешних заимствований мы можем опираться на внешнюю выборку документов, откуда производится заимствование
2. Внутренний. Для поиска внутренний заимствований мы можем использовать только данный нам текст и искать разладки.

Цель эксперимента

Улучшить точность нахождения плагиата в документе при помощи внедрения новых алгоритмов.

Эксперимент

Для вычислительного эксперимента в качестве обучающей выборки использовались данные соревнования **PAN-2018** (корпуса 1-7), а тестирование происходило на данных корпусов 8-10.
Затем надо подобрать оптимальные параметры и запустить с ними



алгоритм.

Модели

1. $G-$ генеративная модель. Для генерации текстов будем использовать *скрытую марковскую цепь*: имея ряд признаков (часто употребляемые слова, синтаксические конструкции и тому подобное) будем стараться угадать, написан ли текст одним автором или несколькими. Для обработки этой модели используем [10], который и будет возвращать нам сгенерированный текст.
2. $D-$ дискриминативная модель, которая представляет собой бинарный классификатор [11], т.е. возможны только два ответа $-0, 1$. Будем ставить 0, если в фрагменте отсутствует плагиат и $1-$ если авторов больше одного. В качестве алгоритма классификации возьмем RandomForestClassifier – он плохо переобучается и устойчив к случайным выбросам и поможет в поиске стилистических разладок.

Предлагаемый алгоритм

Используем generative adversarial networks – генеративная модель порождает тексты в одном авторском стиле, дискриминативная модель - бинарный классификатор.

Результаты эксперимента

По итогам обучения и подбора гиперпараметров получим алгоритм (нейросеть), которая способна находить плагиат в тексте с точностью выше 78

Заключение: результаты, выносимые на защиту

предполагается, что решение этой задачи предлагаемым методом может дать прирост качества по сравнению с типичными методами решениями этой задачи, а также связанных с ней задач кластеризации авторов.