

Style Change Detection

Зуева Надежда

March 2018

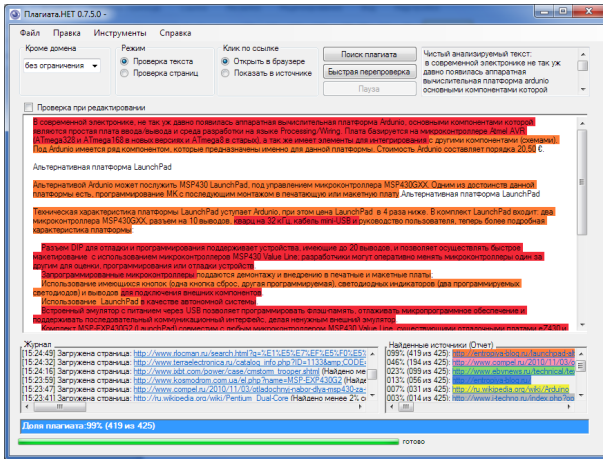
- 1 Написание статей
- 2 Достоверность исторических документов
- 3 Составление профиля автора и восстановление его портрета

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 внутренних заимствований мы можем использовать только данный нам текст и искать разладки.

Существующие методы

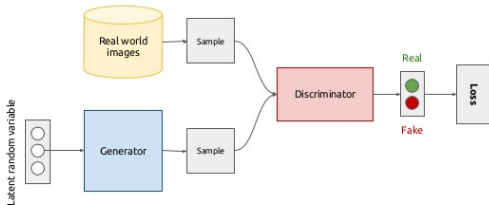
- 1 Построение вектора признаков и поиск отклонений выше некоторой нормы
- 2 Ранние методы, не содержащие машинное обучение, реагирующие на т.н. **стоп-слова**



Предлагаемый алгоритм

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

Generative adversarial networks (conceptual)



5

- ❶ 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.
- ❷ 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.

Ожидаемый результат

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.