

Рецензия на статью
Зуевой Надежды
«Style Change Detection»

В работе рассматривается задача детектирования плагиата в тексте. Для обнаружения плагиата в данной статье предлагается использовать генеративно-сопоставительные нейронные сети. Эксперимент проводится с данными PAN2016, PAN2017 и PAN2018. В качестве критериев качества полученной модели в статье предлагается использовать метрики Precision, Recall, F1-меру и их комбинации.

Рецензируемая статья содержит аннотацию, введение, формальную постановку задачи и описание базового эксперимента. Постановка задачи сформулирована недостаточно четко и не полностью: формально не введена используемая модель (генеративно-сопоставительная сеть), присутствуют одинаковые обозначения для разных понятий (напр., одним и тем же символом D обозначаются и коллекция всех документов, и дискриминативная модель, стр.3, абз.2). Кроме того, в статье присутствуют орфографические неточности, в частности, связанные с математическими обозначениями, что затрудняет чтение статьи (напр., обозначение для совокупности всех заимствованных частей текста, стр.2, абз.5). В целом, раздел “Постановка задачи” структурирован, однако некоторые части дублируют друг друга, что приводит к путанице.

Тема исследования текстов на плагиат довольно актуальна и предложенный в данной работе подход использования генеративно-сопоставительных нейронных сетей обладает актуальностью и новизной. На данный момент в работе есть описание используемого алгоритма, однако отсутствует отчет об эксперименте и полученных результатах.

Однако в статье есть описание базового эксперимента, в котором используется уже существующий алгоритм. В данном разделе демонстрируется отличие заимствованных сегментов от остального текста. На мой взгляд, впоследствии этот раздел можно убрать из статьи (или сократить), поскольку он дублирует результаты, полученные ранее и не несет новизны.

В целом можно сделать вывод, что статья обладает актуальностью и новизной, изложение экспериментов и алгоритма в целом структурировано и соответствует заявленной в аннотации и введении теме, хотя в тексте достаточно много неточностей. Также есть моменты, которые требуют более подробного описания. Если принять во внимание данные замечания, довести до конца эксперимент и получить результаты, на мой взгляд, получится хорошая статья.

Рецензент

Илья Игашов