

УДК: 004.896

Название: Многоагентное моделирование на основе усвоения данных из социальных сетей

Авторы:

Ходорченко М. А., Университет ИТМО, г. Санкт-Петербург;

Контакты: mariyaxod@yandex.ru, +79112142575

Научный руководитель: Бутаков Н. А., Университет ИТМО, г. Санкт-Петербург

Тезис доклада:

Многоагентное моделирование позволяет производить отражение реального мира при помощи взаимодействия агентов. Одной из областей применения подобных моделей может служить воспроизведение общественной динамики, позволяющей проследить за развитием некоторой открытой или закрытой группы людей. Для таких моделей остро встает вопрос проверки их корректности, а именно способа сопоставления с реальными сообществами путем адекватной постановки целевой функции.

В данной работе рассмотрен подход к проверке многоагентной модели, производящей симуляцию сообщества путем моделирования отдельных пользователей и их межличностной коммуникации, для оценивания степени негативной направленности сообщества, а также «негативизации» в динамике. В терминах модели, агентами являются людьми, а взаимодействие осуществляется при помощи генерации сообщений содержащих или не содержащих негатив. Таким образом, для сравнения поведения с реальными сообществами, и приведения информации из текстовой формы в числовую, было необходимо провести подготовительный этап, заключающийся в настройке модели, позволяющей распознавать негатив в сообщениях.

Классификация данных из социальных сетей характеризуется сложностью начального создания тренировочных массивов данных из-за отсутствия размеченных наборов данных и трудоемкости выполнения подобной разметки вручную. Аналогичная ситуация наблюдается и для задачи поиска негативных сообщений. Одним из решений является подход distant supervision, когда происходит перенос знаний из области, где известны метки класса, в область, где их получение затруднено. В данном случае, в качестве домена-донора использовались сайты “Правда сотрудников” и “О Работе”. Предоставленный там формат для написания отзывов включает в себя наличие разделение на позитивные и негативные аспекты, что предоставляет автоматическую разметку на классы. Применение метода distant supervision подразумевает получение зашумленных данных.

В рамках задачи проведения проверки многоагентной модели было произведено сравнение работы нескольких типов архитектур нейронных сетей, а именно BiLSTM, BiQRNN, CNN, а также применение XGBOOST на ELMo эмбедингах. Также был произведен анализ зависимости результатов от различных способов предобработки данных, ограничения словарей, использования подхода извлечения сущностей. Метрикой сравнения алгоритмов была выбрана точность, т. к. целью является наиболее целевое выделение примеров текстов негативной направленности из большой выборки, подаваемой на вход классификатору.

Исходя из полученных результатов, можно заключить, что классификатор может быть использован для извлечения маркеров негативного поведения из текстовых данных с точностью определения 0.894. В работе произведена подготовка нейронной сети, позволяющей настраивать многоагентную модель.

Автор _____

/ __ Ходорченко М. А. __ /

Научный руководитель _____

/ __ Бутаков М. А. __ /