

А.К. Абдуланова¹

¹Международный университет информационных технологий,
г. Алматы, Казахстан

ПОИСК ЛОЖНОЙ ИНФОРМАЦИИ В СТАТЬЯХ С ПОМОЩЬЮ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация. Проблема «поддельных новостей» возникла в последнее время как потенциальная угроза качественной журналистике, и имеет повод к общественной дискуссии в хорошо информированном обществе. Обработка естественного языка является одной из важнейших технологий информационной эпохи. Понимание сложных языковых высказываний также является важной частью искусственного интеллекта. Приложения NLP (Natural Language Processing - обработка естественного языка) повсюду: веб-поиск, реклама, электронные письма, обслуживание клиентов, перевод на язык, радиологические отчеты и т. д. Существует большое количество базовых задач и моделей машинного обучения для приложений NLP. В последнее время глубокие подходы к обучению получили очень высокую производительность во многих различных задачах обработки естественного языка. Они могут решать задачи с помощью отдельных сквозных моделей и не требуют традиционной, специализированной функции.

Ключевые слова: машинное обучение, natural language processing, искусственный интеллект, классификация текста.

• • •

Түйіндеме. «Жалған жаңалықтар» проблемасы сапалы журналистикаға әлеуетті қауіп ретінде соңғы жылдары пайда болды және жақсы хабардар қоғамға пікірталасқа да себеп бар. Табиғи тілді өңдеу – ақпараттық заманның маңызды технологиясының бірі. Сондай-ақ, күрделі пікірді түсіну – жасанды интеллекттің маңызды бір бөлігі. Барлық жерде NLP (Natural Language Processing – табиғи тілді өңдеу) қолданылуда: веб іздеу, жарнама, электрондық хаттар, клиенттерге қызмет көрсету, тілге аудару, радиологиялық есептер және т.б. NLP қосымшалары үшін машиналық оқытудың көптеген базалық есептері мен модельдері бар. Соңғы жылдары оқытуға терең көңіл бөлу табиғи тілді өңдеудің түрлі мәселелерінде үлкен нәтижелерге ие болды. Олар жеке модельдер арқылы мәселелерді шеше алады және мамандандырылған функцияларды талап етпейді.

Түйінді сөздер: машиналық оқыту, табиғи тілдерді өңдеу, жасанды интеллект, мәтінді жіктеу.

Abstract. The problem of “fake news” arose recently as a potential threat to qualitative journalism and a well-informed public debate. The processing of natural language is one of the most important technologies of the information age. Understanding of complex linguistic utterances is also an important part of artificial intelligence. NLP applications (Natural Language Processing) are everywhere: web search, advertising, emails, customer service, translation into language, radiological reports, etc. There are a lot of basic tasks and machine learning models for NLP applications. Recently, deep approaches to learning received very high productivity in many different tasks of natural language processing. They can solve problems using separate end-to-end models and do not require traditional, specialized functions.

Keywords: machine learning, natural language processing, artificial intelligence, text classification.

Введение. В течение последних нескольких месяцев вопрос о «поддельных новостях», определяемый «Нью-Йорк таймс» как «составленные истории, написанные с намерением обмануть» и опубликованные в форматах, подобных тем, что были в традиционных «настоящих» новостях, возникшей как угроза качественной журналистике и хорошо информированного публичного дискурса. В частности, поддельные новости обвинялись в увеличении политической поляризации и партизанского конфликта в Соединенных Штатах во время предвыборной кампании в 2016 г. и в первые дни правления Д.Трампа.

Предположительно, классификатор, который может решить проблему обнаружения положения с высокой точностью, может эффективно использоваться либо как инструмент для людей, работающих для идентификации поддельных новостей (например, получение статей, которые согласны, не согласны и обсуждают заголовки), либо как строительный блок для более сложной системы ИИ, которая попытается определить реальную правдивость новостных сообщений (например, используя достоверные источники для классификации).

За последние несколько лет многие исследовательские усилия в NLP были сосредоточены на применении моделей нейронных сетей с глубокими нейронными сетями для различных задач на основе последовательности. Частым шаблоном для таких задач является принятие двух текстовых последовательностей, их кодирование в какой-то форме, а затем попытка классифицировать их отношения. В

этой задаче был разделен новый состав меченых пар предложений и подходы к их классификации на основе семантических отношений, таких как «влечение» и «противоречие».

Другой общей задачей, основанной на последовательности в NLP, является машинный перевод (МП), который, возможно, является каноническим приложением для недавней работы над структурами «последовательно-последовательно» (seq2seq). Эти модели кодируют одну последовательность слов или токенов, а затем пытаются «декодировать» шаг за шагом соответствующую последовательность выходных токенов. Sutskever от Google сделал одно из первых приложений моделей seq2seq для МП всего несколько лет назад. Другие, включая Bahdanau и Luong расширили работу команды Google, представив новые механизмы, такие как «внимание», что позволяет декодеру модели сфокусироваться на определенных участках закодированной входной последовательности на каждом этапе вывода, чтобы сделать наилучшие предсказания. Поскольку рекуррентные нейронные сети часто являются центральным компонентом моделей NLP, основанных на последовательности, большинство из этих усилий также основываются на единицах «памяти», таких как LSTM (длинная кратковременная память, Hochreiter и Schmidhuber, [1] и GRU (стробированная периодическая единица, Chung).

Феррейра и Влахов [2] использовали набор данных «Emergent» для сравнения слухов о претензиях к новостным статьям, которые ранее были отмечены журналистами с оценкой их правдивости, с целью предсказать позицию статьи в отношении слуха. Эта команда суммировала каждую статью в заголовке и использовала модель логистической регрессии с функциями, представляющими статью, и требовала классифицировать комбинацию статьи и претензии как «для», «против» или «наблюдения» с конечным уровнем точности 73%.

Augenstein предпринял аналогичную задачу обнаружения положения, хотя и в другом наборе данных с несколько более короткими текстовыми строками. Они пытались предсказать, был ли твит «позитивным», «негативным» или «нейтральным» по отношению к короткой теме (например, «Легализация абортов»), которую они обозначили как «цель». Они исследовали несколько моделей, в которых использовалась пара LSTM в разных устройствах. В мо-

дели, которую они называли «независимым кодированием», один LSTM кодировал целевую строку, а другой LSTM кодировал твиты, а конечные векторы скрытого состояния двух LSTM затем пропускались через один слой вперед-вперед и softmax, чтобы сделать предсказание. Окончательным вариантом было «двунаправленное условное кодирование», которое расширило предыдущую модель путем кодирования твита в обоих направлениях, а затем с использованием двух окончательных векторов скрытого состояния для прогнозирования. Результаты этой группы показали, что условное кодирование обеспечило значимое повышение производительности по сравнению с независимой моделью кодирования с двунаправленным условным кодированием, обеспечивающим небольшую дополнительную прибыль.

Задача поиска поддельной новостной статьи. Вызов Fake News был организован в начале 2017 г. для поощрения разработки систем классификации на основе машинного обучения, которые выполняют «обнаружение положения» - то есть то, что конкретный заголовок новостей «соглашается» с «не согласен» с «обсуждает» или не связан с конкретной новостной статьей, чтобы позволить журналистам и другим людям легче находить и исследовать возможные случаи «поддельных новостей». Было разработано несколько моделей, основанных на нейронных сетях, для решения проблемы обнаружения положения, начиная от относительно простых сетей передачи данных, чтобы разработать повторяющиеся модели с вниманием и несколькими словарями.

С появлением поддельных новостей, которые используются для влияния на выборы, определение ложной информации стало важной задачей. Правительства, газеты и платформы социальных сетей прилагают все усилия, чтобы отличить достоверные новости от поддельных новостей. Цель работы - автоматизировать процесс выявления поддельных новостей с помощью машинного обучения и обработки естественного языка. Этот процесс можно разбить на несколько этапов. Первым полезным шагом к идентификации поддельных новостей является понимание того, что другие источники новостей говорят о той же теме. Вот почему проблема фальшивых новостей изначально фокусируется на обнаружении положения. Обнаружение положения включает оценку относительных перспектив двух разных текстовых

фрагментов по той же теме. В частности, задача состоит в том, чтобы оценить позицию заголовка новостей относительно содержания новостной статьи, которая может, но не должна затрагивать ту же тему. Таким образом, относительная позиция каждой пары заголовков-статья должна классифицироваться как не связанная, обсуждающая, соглашающая или не соглашающая.

Открытие несогласованной пары заголовков-статья не обязательно соответствует обнаружению поддельной статьи 1, но это автоматический первый шаг, который мог бы сделать рецензенты для людей осведомленными о несоответствии. В этом случае рецензенты или специализированные алгоритмы могут в конечном итоге решить, какие статьи являются поддельными.

Методы для решения задачи поиска «положения». В этом разделе описаны методы, позволяющие решить задачу поиска «положения».

Сверточные нейронные сети для n-граммов. Вдохновленные успехом базовой линии с символами n-граммов и n-граммами слов, сделана попытка ввести в нашу модель слово n-граммы разного размера. Экспериментировали с конкатенацией слов n-граммов поверх словных вложений, прежде чем кормить их через RNN. Аналогично использовали сверточную нейронную сеть с различными размерами фильтра для генерации представлений n-грамм.

Однако n-граммы не улучшили производительность нашей модели. Опять же, результаты показывают, что введение большего количества параметров, на этот раз в виде сверточных фильтров, может негативно повлиять на производительность нашей модели на этом наборе данных обнаружения положения.

Bag of Words. Некоторые из экспериментов были основаны на совершенно другом подходе, основанном на сумке слов (BoW). Здесь описывается наиболее успешная модель такого рода. Для представления слов использовали 50-мерную версию предварительно подготовленных векторов GloVe [3], используемых в других моделях. Для каждой пары заголовков-корпус. Стоп-слова удаляются как из заголовка, так и из тела. Тело делится на предложения, а средний вектор слова вычисляется для каждого предложения. Соответствующий вектор вычисляется для заголовка. Затем вычисляется сходство косинусов вектора заголовка с каждым

вектором предложения тела и выбирается 3 с наивысшим сходством. Эти векторы, а также вектор заголовка затем объединяются для создания входного вектора классификатора. Объединили глобальные функции с входным вектором. Затем входной вектор подавался в нейронную сеть с одним скрытым слоем 100 ед. ReLU и слоем выхода softmax. Модель BoW работает на удивление хорошо, учитывая ее простоту. [4]

Наивная базовая линия (сходство с Jaccard). Чтобы установить простой базовый уровень производительности, сначала внедрили быстрый счетчик сходства с Jaccard, который сравнивал заголовки с отдельными предложениями от их парной статьи. Найдя максимальные и средние оценки по шкале Jaccard по всем предложениям в статье и выбирая соответствующие пороговые значения, была достигнута 90%-ная точность по связанной/несвязанной задаче, поэтому была надежда на очень высокую точность при переходе к глубокому обучению.

Многоуровневая сеть прямой связи. Наша первая модель на основе NN использовала простые преобразования (усреднение, конкатенацию и т.д.) Для предварительно обученных вложений слов для создания независимых наборов функций для заголовков и статей, которые затем отправлялись через многоуровневую сеть, чтобы получить результат предсказания. Этот метод проходил очень быстро и обеспечивал надежные результаты классификации с общим уровнем точности в диапазоне 90-95% в 4 классах, которые пытались предсказать.

LSTM с независимой условной и двунаправленной условной кодировкой. Вдохновляя подход Аугенштейна и др. [5] было изучено использование нескольких повторяющихся сетевых уровней для кодирования заголовков и статей перед классификацией результирующих векторов состояния с преобразованием softmax. Первая попытка включала отдельные (параллельные) кодировки LSTM заголовка и статьи и последующую классификацию с использованием конечных скрытых состояний каждого кодировщика для прогнозирования. (это то, что Аугенштейн обозначил как «независимую кодировку»). Поскольку словарь по данному набору обучения был относительно небольшим - около 3000 различных типов в заголовках и 24000 типов в статьях - использовались

предварительно подготовленные векторы вставки слов из Стэнфорда GloVe, чтобы дать толчок к захвату семантики маркера. Таким образом, ввод каждого LSTM на каждом шаге представлял собой 50-мерное векторное представление текущего токена, основанное на GloVe, но также прошедшее обучение в данной модели, чтобы поймать любые корректировки, которые могут быть специфическими для задачи классификации.

Затем перешли к условному кодированию, которое включает повторяющиеся ячейки, размещенные в последовательности, а не параллельно. Сначала был отправлен текст заголовка через слой LSTM (назовем эту LSTMHeadline), а затем инициализирован другой слой LSTM (назовём эту LSTMА-страницу) конечным скрытым вектором состояния LSTMHeadline.

Затем окончательное предсказание было выполнено с использованием только конечного скрытого состояния LSTMА-статьи, хотя, конечно, концептуальная цель заключалась в том, что это состояние также фиксирует информацию о заголовке и его отношении к статье на основе ее инициализации. Наконец, была попробована двунаправленная условная кодировка, состоящая из 4 разных слоев LSTM.

LSTMHeadline-Forward подключается к LSTMА-Forward-Forward, как и в предыдущей модели, в то время как LSTMHeadline-Backward и LSTMА-Article-Backward имеют одинаковые отношения, но их входы поставляются в обратном порядке. Затем конечные состояния LSTMА-Прямой и LSTMА-Article-Backward затем усредняются и подаются в конечный слой предсказания softmax.

Последовательная повторяющаяся модель с вниманием. Все предыдущие модели были, по существу, созданы на заказ в Python и Tensorflow, хотя использовались некоторые полезные строительные блоки из библиотеки Tensorflow, такие как класс BasicLSTMCell. Для окончательной модели было решено поэкспериментировать с некоторыми из более богатых библиотечных кодов, доступных для Tensorflow.

Выводы. В частности, было адаптировано учебное пособие по переводам нейронных машин (NMT) последовательности к последовательности, чтобы создать внимательную модель seq2seq, используя функцию «embedding_attention_seq2seq» от Tensorflow.

Эта модель позволила легко использовать LSTM или GRU, а также обеспечила преимущество добавления уровня внимания поверх условного кодирования, которое является стандартным в моделях seq2seq.

Список литературы

1 *Pomerleau D. and Rao D.*, "Fake news challenge." 2016., [Электронный ресурс]: <http://www.fakenewschallenge.org>

2 *Ferreira W. and Vlachos A.*, "Emergent: a novel data-set for stance classification," in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, ACL, 2016.

3 *Pennington J., Socher R., and Manning C.D.*, "Glove: Global vectors for word representation," in Empirical Methods in Natural Language Processing (EMNLP), 2014.- pp. 1532–1543,

4 *Mohammad S.M., Kiritchenko S., Sobhani P., Zhu X., Cherry C.*, "Detecting stance in tweets," Proceedings of SemEval, 2016.- vol. 16.

5 *Augenstein I., Rocktaschel T., Vlachos A., Bontcheva K.*, "Stance detection with bidirectional conditional encoding," 2016.- 1606.05464

Абдуланова А.К., e-mail: abdulanova@gmail.com