

МЕТОД ПРИМЕНЕНИЯ НЕЙРОННЫХ СЕТЕЙ BERT-BILSTM-ATTENTION ДЛЯ ОПРЕДЕЛЕНИЯ ЭМОЦИОНАЛЬНОГО ОТНОШЕНИЯ АВТОРА К ТЕКСТУ

Инь Суэзюнь

Московский государственный технический
университет им. Н.Э. Баумана (национальный
исследовательский университет)
1535861547@qq.com

Афанасьев Геннадий Иванович

К.т.н., доцент, Московский государственный
технический университет им. Н.Э. Баумана
(национальный исследовательский университет)
gaipcs@bmstu.ru

Калистратов Алексей Павлович

Ассистент, Московский государственный технический
университет им. Н.Э. Баумана
(национальный исследовательский университет)
akalistratov@gmail.com

**BERT-BILSTM-ATTENTION NEURAL
NETWORKS APPLICATION METHOD
TO DETERMINE THE AUTHOR'S
EMOTIONAL ATTITUDE TO THE TEXT**

S. Yin

G. Afanasyev

A. Kalistratov

Summary: In this paper, a text sentiment analysis method based on the BERT-BiLSTM-Attention model is proposed. During preparation, the text is encoded using the BERT model to obtain a semantic representation of the text. Then, the encoded text is uploaded into the BiLSTM model to obtain deeper semantic information. Finally, the output of the BiLSTM model is weighted and averaged by the Attention mechanism to obtain the final sentiment analysis results. Experimenting on the IMDB dataset, the model proposed in this paper achieved an accuracy rate of 90%, providing better performance than other deep learning-based sentiment analysis methods.

Keywords: sentiment analysis, BERT-BiLSTM-Attention, Attention mechanism, deep learning, neural networks.

Аннотация. В данной работе предлагается метод анализа текста на предмет его тональности, основанный на модели глубокого обучения нейронных сетей BERT-BiLSTM-Attention. Сначала текст кодируется с помощью модели BERT для получения его семантического представления. Затем закодированный текст поступает в модель BiLSTM для обработки этого представления в виде вектора. После этого выход модели BiLSTM взвешивается и усредняется алгоритмом Attention для получения окончательных результатов анализа настроения. Экспериментируя на наборе данных IMDB, модель, предложенная в данной работе, достигла точности 90 %, обеспечив лучшую производительность по сравнению с другими методами текстового анализа на основе глубокого обучения.

Ключевые слова: анализ тональности текста; BERT-BiLSTM-Attention; механизм Attention, глубокое обучение, нейронные сети.

Введение

Анализ тональности текста — важная задача в направлении обработки естественного языка, также известная как распознавание настроений, поиск настроений и т.д. Суть задачи — это автоматический анализ и распознавание человеческих эмоций, выраженных в тексте, речи и т.д. Потребность в анализе настроений варьируется от частных лиц до крупных организаций и правительств. Отдельный клиент сначала проверяет рейтинги продукта и мнения других покупателей, прежде чем принять решение о покупке, а бизнес-организации используют инструменты анализа настроений, чтобы понять чувства клиентов. Было предпринято несколько попыток применить анализ настроений к отзывам клиентов. Таким образом, было доказано, что анализ отзывов клиентов способствует развитию соответствующего рынка и повышению доверия клиентов. Кроме того, правительства анализируют чувства общественности по поводу таких актуальных тем, как выборы и их

политика. Недавним примером является предсказание президентских выборов в США в 2016 году. Кроме того, анализ настроений может быть использован для расширения возможностей рекомендательных систем, где можно идентифицировать интересы пользователей. Анализ тональности используется в различных веб-приложениях, например, для анализа мнений в социальных сетях, анализа отзывов о товарах, для мониторинга брендов и так далее [1–7].

Модель BERT-BiLSTM-Attention — это модель глубокого обучения для задач классификации текстов, особенно анализа настроений. Она объединяет модель предварительного обучения BERT, двунаправленную модель LSTM и механизм внимания для улучшения производительности и обобщения модели [8].

Эта модель хорошо справляется с задачами анализа настроений, особенно при работе с длинными текстами, например может лучше аналогов передавать контекст-

ную и семантическую информацию текста [9]. Однако, данная модель для работы требует больших вычислительных затрат, поскольку для ее обучения требуется много времени и большие вычислительные ресурсы. Кроме того, для обучения модели требуется большой объем размеченных данных, иначе она может не достичь оптимальной производительности [10].

Описание модели

BERT использует для кодирования текста двунаправленную структуру, т.к., такая структура кодирования демонстрирует способность модели описывать определенную семантику других слов в контексте через семантические отношения контекста при обработке конкретного слова [11].

Промышленное и академическое сообщество обратило внимание на анализ настроений, чтобы помочь в принятии решений. Следовательно, исследователи предложили большое количество подходов для того, чтобы удовлетворить потребность в анализе настроений. Доказано, что предложенные традиционные подходы дают хорошие результаты при правильном проектировании функций. Таким образом, наиболее часто используемые функции этих подходов анализируются: части речи, позиция термина, слова-мнения и предложения, отрицание, присутствие и частотность термина и синтаксическая зависимость. Традиционные подходы к анализу настроений подразделяются на две категории: подходы, основанные на лексике, и подходы к машинному обучению. Подходы, основанные на лексиконе, являются разновидностями традиционных подходов к анализу настроений, которые используют предварительно скомпилированные словари настроений, содержащие разные слова и их полярность, для классификации данного слова по классам положительных или отрицательных настроений. В исследованиях приводится подробное описание этих подходов. В связи с популярностью лексиконов были проведены обширные исследования по анализу настроений на основе лексиконов. Эти подходы не требуют наличия обучающего набора данных. Однако построение словаря настроений для современных пользовательских неструктурированных данных является сложной задачей. Следовательно, подходы к машинному обучению помогают облегчить проблему. Подходы к машинному обучению — это другие традиционные методы анализа настроений, которые основаны на алгоритмах машинного обучения для классификации слов по соответствующим им меткам настроений. Основным преимуществом подходов к машинному обучению является их способность к обучению представлениям. Алгоритмам машинного обучения требуется обучающий набор данных, который помогает автоматизировать классификатор, и тестовый набор данных, используемый для проверки работоспособности классификатора.

Таким образом, подходы машинного обучения являются предпочтительными для анализа настроений из-за их способности обрабатывать большие объемы данных по сравнению с подходами, основанными на лексике. Однако в случае отсутствия наборов данных, аннотированных человеком, большинство людей предпочитают использовать подходы, основанные на лексике. Таким образом, обширные исследования было сделано для анализа настроений с использованием подходов машинного обучения. Что касается хороших результатов анализа настроений, то два традиционных подхода к анализу настроений могут быть объединены, чтобы получить преимущества каждого подхода. Таким образом, недавние достижения являются результатами сочетания двух традиционных подходов.

В данной работе мы изучаем то, что текст преобразуется BERT в числовую матрицу, где каждый столбец представляет идентифицированный признак, а каждая строка — конкретный комментарий. Эта матрица используется в качестве входных данных для BERT, модель обучается с помощью двух стратегий обучения: MLM и предсказание следующего предложения. Модель BERT показана на рисунке 1 [12].

После завершения обучения эти данные передаются в модель BiLSTM. Двунаправленная нейронная сеть с долговременной памятью (BiLSTM) — это двойная модель, которая объединяет прямую LSTM и обратную LSTM для лучшей обработки двунаправленных семантических признаков. Таким образом, модель BiLSTM состоит из двух инвертированных компонентных моделей LSTM, с выходом из двух совместных состояний принятия решений [13]. Схема работы этой модели представлена на рисунке 2.

Обновление LSTM можно описать следующим образом:

$$h_t^1 = H(W_{xh_t^1}x_t + W_{h^1h^1}h_{t-1}^1 + b_{h^1}) \quad (1)$$

Формулу обновления LSTM можно описать следующим образом:

$$h_t^2 = H(W_{xh_t^2}x_t + W_{h^2h^2}h_{t-1}^2 + b_{h^2}) \quad (2)$$

Выходной слой LSTM можно описать следующим образом:

$$y_t = W_{h^1y}h_t^1 + W_{h^2y}h_t^2 + b_y \quad (3)$$

В анализе тональности внедрение механизма Attention может лучше сработать, если стоит задача изучить «межсловные» зависимости и усилить фокус на ключевых словах — в зависимости от порядка, важность слов в тексте меняется, т.е., и веса признаков различаются [14–16]. Наконец, вектор признаков h по-

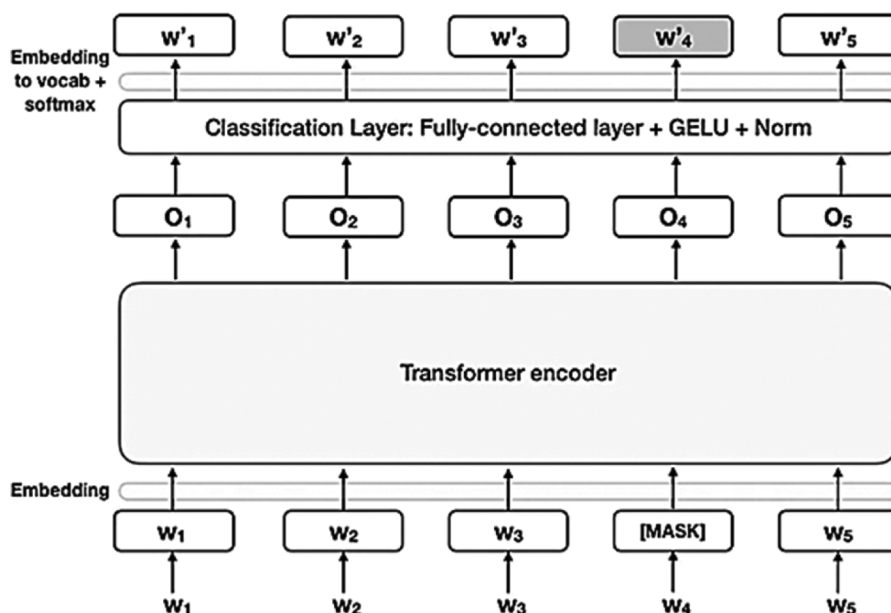


Рис. 1. Структура модели BERT

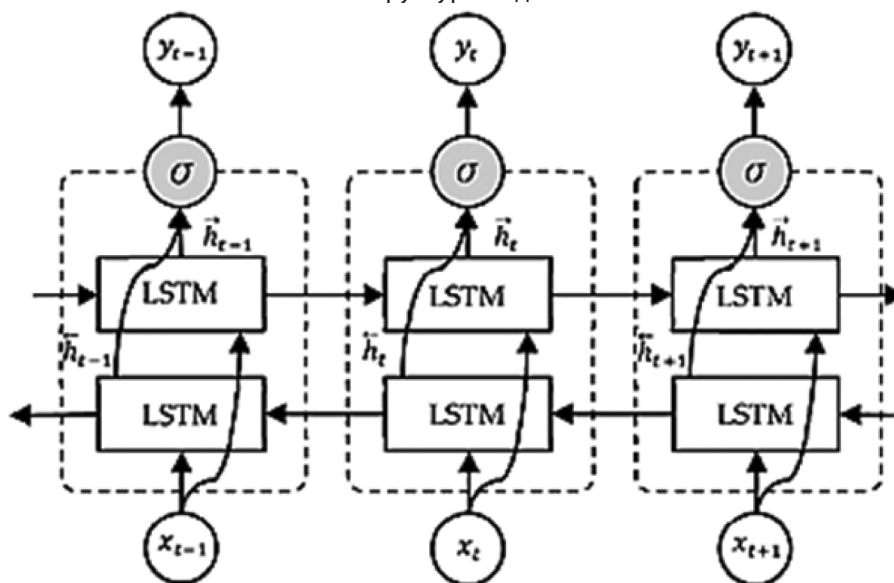


Рис. 2. Структура модели BiLSTM

ступает на плотный слой модели, где его размерность уменьшается по формуле 2. Предсказание тональности определяется по формуле 1, где W и b обозначают вектор параметров и смещение, соответственно.

Проведение эксперимента

Выбранная модель BERT-BiLSTM-Attention используется для анализа тональности текстов длиннее 140 символов, при этом в качестве обучающих и тестовых данных используются тексты из базы соревнований Kaggle. Все наборы данных были разделены на обучающий набор, проверочный набор и тестовый набор, с пропорцией 80 %, 10 % и 10 % от общего набора данных соответственно.

В модели были внесены предварительные настройки, основанные на опыте других исследователей [8, 9]. Например, для настройки BERT и других «обычных» моделей был применен оптимизатор Adam с $K = 0.00002$, $N_{party} = 32$, $N_{training} = 3$.

После модели были обучены и проверялись на трех наборах данных: отзывы на фильмы, отзывы на рестораны, описание фильмов с IMDB. Результаты экспериментов и сравнение точности модели с добавлением Attention приведены в таблице 1. Сравнение производилось с известными по другим статьям моделям BERT, BiLSTM и BERT-BiLSTM.

Таблица 1.

Сравнение точности моделей

Модели	Фильмы	Рестораны	IMDB
BERT	0.859	0.776	0.831
BiLSTM	0.825	0.720	0.828
BERT-BiLSTM	0.870	0.781	0.896
BERT-BiLSTM-Attention	0.890	0.812	0.903

После экспериментов можно сделать следующие выводы:

- Модель BERT относительно плохо работает на двух других наборах данных.
- Модель BiLSTM плохо работает на всех наборах данных.
- Модель BERT-BiLSTM работает лучше двух предыдущих на всех наборах данных.

В итоге, модель BERT-BiLSTM-Attention показала лучшие результаты на всех наборах данных, что говорит о целесообразности усиления двух моделей за счет третьей для задачи распознавания тональности текстов длиннее 140 символов.

Заключение

В данной работе была рассмотрена многосоставная модель на основе нескольких уже проверенных различными исследователями моделей глубокого обучения. Был найден способ повышения точности, который показал хоть и небольшой, но выигрыш относительно других методов. В последние годы было предложено несколько подходов к анализу настроений, которые были впервые внедрены в Китае и других странах. Эти традиционные подходы позволили достичь хороших результатов, но разработка функций, на которые они опираются, является утомительной задачей. Позже исследователи поняли, что поиск настроений для современных пользовательских данных требует глубокого понимания, и для того, чтобы справиться с этим, необходимы эффективные методы. Таким образом, подходы к глубокому обучению эволюционировали как эффективные методы благодаря их способности изучать текст без ручной разработки функций. Доказано, что подходы глубокого обучения превосходят традиционные методы анализа настроений. Таким образом, мы рассматриваем последние модели глубокого обучения, которые были предложены для различных задач анализа настроений. В будущем авторы планируют провести еще несколько экспериментов для подтверждения этого вывода.

ЛИТЕРАТУРА

1. Афанасьев Г.И., Афанасьев А.Г., Бурмистрова М.В. Тэт В.Я.С. Исследование методов машинного обучения для прогнозирования эффективных бизнес-решений в системах электронной коммерции // E-Scio. 2022. № 11 (74). С. 59–72.
2. Афанасьев Г.И., Афанасьев А.Г., Зо Х.А. Анализ возможностей применения компьютерного зрения в рекомендательных интернет-системах // E-Scio. 2022. № 11 (74). С. 73–82.
3. Афанасьев Г.И., Гой А.В., Ковалева Н.А. Анализ и сравнение методов Process Mining // Аспирант и соискатель. 2020. № 1 (115). С. 195–199.
4. Ч. Чжан, Афанасьев Г.И. Основные технологии и перспективы эволюции персонализированных рекомендательных систем // E-Scio. 2022. № 4(67). С. 309–320.
5. Ч. Чжан, Афанасьев Г.И. Применение моделей глубокого обучения в области рекомендательных систем, основанных на содержании // Искусственный интеллект в автоматизированных системах управления и обработки данных. Сборник статей Всероссийской научной конференции. В 2-х томах. Москва: МГТУ им. Н.Э. Баумана. 2022. Т.1. С. 278–284.
6. Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. // aclanthology.org. <https://aclanthology.org/N19-1423/>. 2019. DOI:10.18653/v1/N19-1423.
7. Hochreiter, S., & Schmidhuber, J. (). Long short-term memory. // Neural computation. 1997. 9(8). P.1735–1780.
8. Shaukat Z., Zulfqar A.A., Xiao C., Azeem M., & Mahmood T. // Sentiment analysis on IMDB using lexicon and neural networks. // SN Applied Sciences, 2020. 2. P.1–10.
9. Lee, Lung-Hao, et al. «NCUEE at MEDIQA 2019: medical text inference using ensemble BERT-BiLSTM-Attention model». // Proceedings of the 18th BioNLP workshop and shared task. 2019. P.528–532.
10. Liu, B. Sentiment analysis and opinion mining. // Synthesis lectures on human language technologies. 2012. 5(1). P.1–167.
11. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., & Potts, C. Learning word vectors for sentiment analysis. // In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1. 2011. P.142–150.
12. Pennington, J., Socher, R., & Manning, C. D. Glove: Global vectors for word representation. // In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014. P.1532–1543.
13. Zhang, X., Zhao, J., & LeCun, Y. Character-level convolutional networks for text classification. // In Advances in neural information processing systems. 2015. P. 649–657.
14. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2018). Attention-based bidirectional long short-term memory networks for relation classification. // In Proceedings of the 27th international conference on computational linguistics. 2018. P. 3249–3259.
15. Zhou, Y., Xu, C., & Huang, Y. End-to-end learning of semantic role labeling using recurrent neural networks. // In Proceedings of the 54th annual meeting of the association for computational linguistics. volume 1. 2018. P. 1127–1137.
16. Labeled Faces in the Wild // CS.UMASS.EDU. 9 January 2018. URL: <http://vis-www.cs.umass.edu/lfw/> (date of access:20.03.2022).

© Инь Сузюнь (1535861547@qq.com); Афанасьев Геннадий Иванович (gaipecs@bmsu.ru);

Калистратов Алексей Павлович (akalistratov@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»