MODELING AND ANALYSIS OF INFORMATION SYSTEMS, VOL. 31, NO. 1, 2024
journal homepage: www.mais-journal.ru

ARTIFICIAL INTELLIGENCE

Application of Deep Neural Networks for Automatic Irony Detection in Russian Texts

M. A. Kosterin[1], I. V. Paramonov[1]

DOI: 10.18255/1818-1015-2024-1-90-101

[1]P.G. Demidov Yaroslavl State University, Yaroslavl, Russia

MSC2020: 68T50
Research article
Full text in Russian

Received February 15, 2024
Revised February 23, 2024
Accepted February 28, 2024

The paper examines automatic methods for classifying Russian-language sentences into two classes: ironic and non-ironic. The discussed methods can be divided into three categories: classifiers based on language model embeddings, classifiers using sentiment information, and classifiers with embeddings trained to detect irony. The components of classifiers are neural networks such as BERT, RoBERTa, BiLSTM, CNN, as well as an attention mechanism and fully connected layers. The irony detection experiments were carried out using two corpora of Russian sentences: the first corpus is composed of journalistic texts from the OpenCorpora open corpus, the second corpus is an extension of the first one and is supplemented with ironic sentences from the Wiktionary resource.

The best results were demonstrated by a group of classifiers based on embeddings of language models with the maximum F-measure of 0.84, achieved by a combination of RoBERTa, BiLSTM, an attention mechanism and a pair of fully connected layers in experiments on the extended corpus. In general, using the extended corpus produced results that were 2–5 % higher than those of the basic corpus. The achieved results are the best for the problem under consideration in the case of the Russian language and are comparable to the best one for English.

Keywords: irony detection; sarcasm detection; neural network-based classifier; deep learning; natural language processing; BERT

INFORMATION ABOUT THE AUTHORS

Kosterin, Maksim A.	ORCID iD: 0000-0002-8298-3156. E-mail: makcost@gmail.com Post-graduate student
Paramonov, Ilya V. (corresponding author)	ORCID iD: 0000-0003-3984-8423. E-mail: ilya.paramonov@fruct.org PhD, associate professor

Funding: Russian Science Foundation (Project no. 23-21-00495).

For citation: M. A. Kosterin and I. V. Paramonov, “Application of deep neural networks for automatic irony detection in Russian texts”, *Modeling and Analysis of Information Systems*, vol. 31, no. 1, pp. 90–101, 2024. DOI: 10.18255/1818-1015-2024-1-90-101.

МОДЕЛИРОВАНИЕ И АНАЛИЗ ИНФОРМАЦИОННЫХ СИСТЕМ, ТОМ 31, № 1, 2024
сайт журнала: www.mais-journal.ru

Применение глубоких нейронных сетей для автоматического определения иронии в русскоязычных текстах

М. А. Костерин[1], И. В. Парамонов[1]

DOI: 10.18255/1818-1015-2024-1-90-101

[1]Ярославский государственный университет им. П.Г. Демидова, Ярославль, Россия

УДК 004.912
Научная статья
Полный текст на русском языке

Получена 15 февраля 2024 г.
После доработки 23 февраля 2024 г.
Принята к публикации 28 февраля 2024 г.

В работе исследуются автоматические методы классификации русскоязычных предложений на два класса: содержащие и не содержащие ироничный посыл. Рассматриваемые методы могут быть разделены на три категории: классификаторы на основе эмбеддингов языковых моделей, классификаторы с использованием информации о тональности и классификаторы с обучением эмбеддингов обнаружению иронии. Составными элементами классификаторов являются нейронные сети, такие как BERT, RoBERTa, BiLSTM, CNN, а также механизм внимания и полносвязные слои. Эксперименты по обнаружению иронии проводились с использованием двух корпусов русскоязычных предложений: первый корпус составлен из публицистических текстов из открытого корпуса OpenCorpora, второй корпус является расширением первого и дополнен ироничными предложениями с ресурса Wiktionary.

Лучшие результаты продемонстрировала группа классификаторов на основе чистых эмбеддингов языковых моделей с максимальным значением F-меры 0.84, достигнутым связкой из RoBERTa, BiLSTM, механизма внимания и пары полносвязных слоев в ходе экспериментов на расширенном корпусе. В целом использование расширенного корпуса давало результаты на 2–5 % выше результатов на базовом корпусе. Достигнутые результаты являются лучшими для рассматриваемой задачи в случае русского языка и сравнимы с лучшими для английского.

Ключевые слова: обнаружение иронии; обнаружение сарказма; нейросетевой классификатор; глубокое обучение; обработка естественного языка; BERT

ИНФОРМАЦИЯ ОБ АВТОРАХ

Костерин, Максим Алексеевич	ORCID iD: 0000-0002-8298-3156. E-mail: makcost@gmail.com Аспирант
Парамонов, Илья Вячеславович (автор для корреспонденции)	ORCID iD: 0000-0003-3984-8423. E-mail: ilya.paramonov@fruct.org Канд. физ.-мат. наук, доцент

Финансирование: Российский научный фонд (проект № 23-21-00495).

Для цитирования: M. A. Kosterin and I. V. Paramonov, “Application of deep neural networks for automatic irony detection in Russian texts”, *Modeling and Analysis of Information Systems*, vol. 31, no. 1, pp. 90–101, 2024. DOI: 10.18255/1818-1015-2024-1-90-101.

Введение

Автоматическое обнаружение иронии и сарказма — одна из задач обработки естественного языка, направленная на выявление в отдельных предложениях или текстах содержания, имеющего иронический или саркастический посыл. Для русского языка эта задача на настоящий момент остается малоисследованной. Вызвано это в первую очередь недостатком корпусов текстов на русском языке, размеченных по наличию и отсутствию иронии и сарказма. Тем не менее эта задача является актуальной в такой области, как анализ общественного мнения, также её решение может быть полезно для других задач автоматической обработки текста, например, классификации по тональности.

Данная статья развивает рассмотренные в работе [1] автоматизированные методы классификации русскоязычных предложений на два класса: содержащие и не содержащие ироничный посыл. Сарказм в обеих работах предполагается частным случаем иронии. В исходной работе исследовались только базовые нейросетевые модели, такие как BERT и BiLSTM, а также алгоритмы классического машинного обучения: логистическая регрессия, методы опорных векторов и случайного леса. Целью текущей работы является изучение возможностей более сложных моделей и методов, использующих связки классификаторов и различные подходы к классификации в рамках одной сложной архитектуры, для повышения качества результатов.

В качестве основных подходов в текущей работе используются методы с использованием эмбеддингов чистых языковых моделей (Language Modeling, LM), а также эмбеддингов языковых моделей, адаптированных под предметную область задач классификации по тональности (Sentiment Analysis, SA) и обнаружения иронии (Irony Detection, ID). Здесь LM-эмбеддингами называются эмбеддинги предпоследнего внутреннего слоя BERT, который обучался задаче маскированного моделирования языка (MLM), SA-эмбеддингами — эмбеддинги предпоследнего внутреннего слоя BERT, обученного классифицировать тональность, а под ID-эмбеддингами понимаются эмбеддинги предпоследнего внутреннего слоя BERT, который дополнительно к предобучению MLM обучался обнаружению иронии. В качестве структурных элементов классификаторов используются нейронные сети BERT, RoBERTa, BiLSTM, CNN и механизм внимания (Attention).

Следует отметить, что в отличие от подавляющего большинства исследований, встречающихся в литературе, в данной работе для экспериментов используются не короткие тексты из социальных сетей Twitter, Reddit и им подобным, а тексты из новостных и аналитических статей и записи блогов. Это обеспечивает дополнительную новизну исследованию, а также приводит к некоторому усложнению задачи, поскольку в подобных текстах используется значительно более широкий спектр средств выражения иронии.

Статья построена следующим образом. В разделе 1 приведён обзор работ по тематике автоматического определения иронии и сарказма в текстах. Детальное описание используемых в проведённом исследовании методов и моделей приводится в разделе 2. Раздел 3 описывает корпуса, используемые для проведения экспериментов. В разделе 4 производится оценка предложенных методов и моделей на двух корпусах русскоязычных предложений. В заключении подведены итоги работы.

1. Обзор связанных работ

Наиболее распространенным подходом к обнаружению иронии является применение векторов характеристик для кодирования текстовых последовательностей с последующей бинарной классификацией этих векторов. Векторы характеристик обычно строятся с использованием эмбеддингов, генерируемых такими инструментами, как Word2Vec, GloVe, BERT и т. д.

Использование базовых подходов для обнаружения сарказма в англоязычном корпусе, собранном из текстов социальной сети Reddit, демонстрируется в работе [2]. В ней приводится сравнение

классификаторов, построенных с использованием методов машинного обучения и нейронных сетей BERT и BiLSTM, при анализе комментариев пользователей. Лучшим среди методов машинного обучения оказался метод гребневой регрессии, позволяющий получить F-меру 0.68. В целом же наилучший результат демонстрирует нейронная сеть BERT, достигшая F-меры 0.72. Схожее сравнение базовых подходов для русского языка из предшествующей работы [1] показало близкие результаты: BERT также занял первое место с F-мерой 0.76, а среди методом машинного обучения лучше всего себя показала логистическая регрессия с F-мерой 0.68.

Более совершенный подход с применение эмбеддингов RoBERTa в связке с BiLSTM, аналог которого применяется и в текущей работе, для англоязычных текстов рассматривается в работе [3]. В ней предобученные эмбеддинги RoBERTa являются входными данными для слоя BiLSTM, а также конкатенируются с его выходными данными в один вектор, используемый в классификаторе. Предложенный подход показывает F-меру 0.80, 0.78 или 0.90 для корпусов текстов Irony/SemVal-2018-Task 3.A [4], Reddit SARC2.0 politics [5] и Riloff sarcastic dataset [6] соотвественно.

В работе [7] авторами исследуется применение подхода трансферного обучения для решения задачи обнаружения иронии. В ней используются две нейронный сети BiLSTM: одна из них сперва обучается задаче классификации тональности на корпусе англоязычных твитов; далее веса первой BiLSTM замораживаются и эти же твиты подаются на вход уже обеим BiLSTM; наконец, выходные данные обеих BiLSTM объединяются в один вектор и передаются в последовательно расположенные слой внимания и полносвязный слой, выполняющий функцию классификатора для задачи обнаружения иронии. В данной структуре первая BiLSTM, обученная классифицировать тональность, предназначена для выделения тональных характеристик, которые дополняют семантические характеристики, извлекаемые второй BiLSTM. С предложенным подходом в работе достигается F-мера 0.68 на корпусе Irony/SemVal-2018-Task 3.A [4] и 0.78 на корпусе Riloff sarcastic dataset [6]

В работе [8] предлагается метод обнаружения сарказма в текстовых онлайн-дискуссиях с использованием контекста — характеристик пользователя, оставившего комментарий, включающие его стилометрические характеристики, личностные характеристики и характеристики сообщений форума, в рамках которого ведется дискуссия. Моделирование контекста производится путем использования CNN для извлечения высокоуровневых характеристик из истории сообщений пользователя и форума. Позже эти характеристики добавляются к содержательным характеристикам классифицируемого сообщения, также извлеченным с помощью CNN, для обучения модели решению задачи обнаружению сарказма. В качестве корпуса данных в этой работе применяется Reddit SARC2.0 [5]. Предложенная модель без использования личностных характеристик показывает F-меру 0.66, а при их использовании этот показатель повышается до 0.77.

Проблема создания корпуса русскоязычных текстов для обнаружения сарказма освещается в работе [9]. Согласно полученным в ней данным, хэштеги, являющиеся маркерами сарказма или иронии, не используются в достаточной степени русскоязычной аудиторией Twitter. Отмечается, что хэштег #ирония в основном употребляется в качестве описания изображений или шуток вместе с хэштегами #шутка и #смех. В тоже время хэштег #сарказм действительно применяется для обозначения сарказма, однако его релевантное употребление недостаточно распространено, чтобы использовать его для составления корпуса.

В работе [10] задача обнаружения сарказма в текстах на русском языке решается посредством решения задачи классификации по тональности. Авторами используется корпус предложений Twitter, который сперва разделяется на три класса тональности (положительный, отрицательный, нейтральный) нейронной сетью RuBERT, обученной решению задачи классификации тональности. Далее на размеченном корпусе обучается модель для обнаружения сарказма следующим образом: с использованием словаря тональности слов подсчитывается количество положительных и отрицательных слов в предложении; если предложение содержит слова с отрицательной тональностью

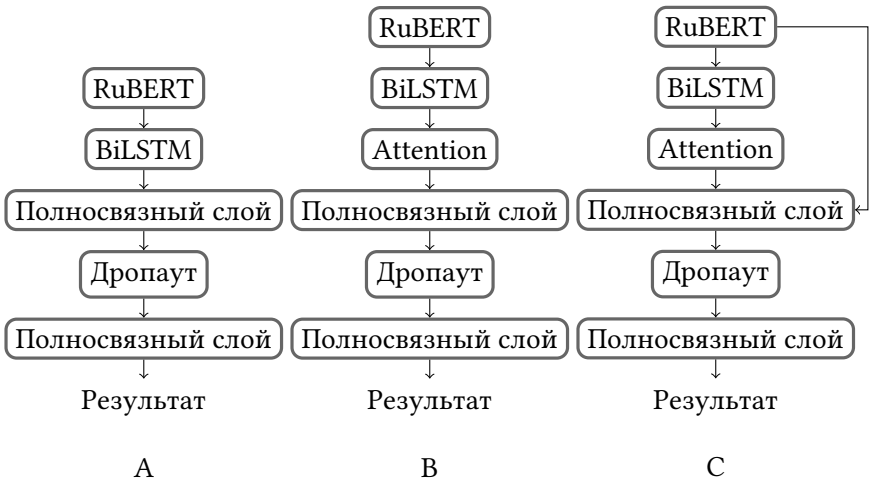

Fig. 1. RuBERT (LM) — BiLSTM model

Рис. 1. Модель RuBERT (LM) — BiLSTM

и при этом количество слов с положительной тональностью превышает количество слов с отрицательной, но класс тональности данного предложения является отрицательным, то в таком предложении отмечается наличие сарказма. Наличие сарказма также отмечается и в противоположном случае, когда количество отрицательно окрашенных слов превышает количество положительных, а класс тональности предложения является положительным. Таким образом задача обнаружения сарказма в данной работе решается без выполнения разметки корпуса по наличию или отсутствию сарказма, а использует только информацию о тональности. F-мера предложенного подхода составляет 0.69.

В приведенных работах описывается только часть подходов, используемых для решения задачи обнаружения иронии и сарказма, а более подробный обзор области приведен в работе [11]. Рассмотренные же в данном разделе работы выделяются в том плане, что используемые ими подходы, модели или данные представляют интерес в рамках исследования данной работы. В целом можно отметить, что задача обнаружения иронии является менее исследованной, чем многие другие задачи обработки естественных языков, например, классификации тональности, вопросно-ответные системы или машинный перевод. Особенно это актуально для русского языка, где количество исследований по данной теме минимально, а проведение экспериментов сопровождается определенными трудностями в плане сбора экспериментальных данных.

2. Используемые методы и модели

2.1. Классификаторы на основе эмбеддингов

В данную группу входят классификаторы, которые оперируют эмбеддингами токенов, генерируемыми моделью RuBERT [12]. Для каждого подаваемого на вход текстового токена RuBERT производит соответствующий ему числовой вектор. В дальнейшем обучение классификатора происходит с использованием последовательностей этих векторов.

Наиболее простым классификатором из данной группы является связка из RuBERT с BiLSTM, дополненная парой полносвязных слоев (рис. 1A). В данной связке RuBERT отвечает за создание контекстуализированных эмбеддингов, BiLSTM — за моделирование контекста входной последовательности, первый полносвязный слой с функцией активации ReLU — за классификацию, а второй полносвязный слой с функцией активации softmax выполняет роль выходного слоя, преобразуя вектор результатов классификатора в значения предсказания. Два полносвязных слоя соединены слоем дропаута для предотвращения переобучения.

Следующая модель в группе дополнена слоем внимания, расположенным между BiLSTM и первым полносвязным слоем (рис. 1B). Задачей слоя внимания является назначение весов токенам входной последовательности в зависимости от их важности при определении того, является ли предложение ироничным или нет. Конструкции, которые являются наиболее явным признаком наличия иронии, будут иметь большие веса и наоборот.

Следующая модель также расширяет предыдущую, в этот раз добавляя значения эмбеддингов токенов к выходным значениям слоя внимания (рис. 1C). Идея данного подхода заключается в том, чтобы вместе с моделью входной последовательности в слой классификатора передать и изначальные сведения об контексте употребления токенов, тем самым увеличив количество доступной информации о токене. Эмбеддинги токенов и вектор внимания каждого предложения конкатенируются поэлементно между собой и передаются на вход первому полносвязному слою. Для этой модели также была проверена ее вариация, в которой вместо RuBERT для формирования эмбеддингов использовалась модель RoBERTa [13].

Следующие три модели из этой группы помимо эмбеддингов учитывают также и характеристики предложения, потенциально указывающие на наличие иронии в нем (рис. 2A). В первой модели в качестве характеристик выступают только определенные знаки препинания («?», «!» и «. . . »), кавычки и круглые скобки как части смайлов. Во второй модели к признакам добавляется информация о присутствующих в предложении частях речи и биграммах, а в третьей — информация о наличии в предложении последовательностей токенов, входящих в список ироничных выражений, который был собран с использованием русскоязычной версии Интернет-ресурса Wiktionary[1]. В качестве ироничных выражений собирались фразы, принадлежащие к категории «Ироничные выражения/ru». В результате получился список из 1144 крылатых выражений, жаргонизмов, фразеологизмов и прочих фраз и слов, употребляемых с целью высмеять или оскорбить. Информация о характеристиках предложений преобразовывается в числовой вектор, который конкатенируется с вектором внимания, после чего результат подается на вход первому полносвязному слою. Таким образом классифицируемый вектор содержит как контекстуализированную модель предложения, так и информацию о некоторых характеристиках текста, потенциально сигнализирующих о наличии иронии.

Последний классификатор из данной группы (рис. 2B) использует два последовательно идущих слоя CNN с размером ядра 3 для выделения последовательностей высокоуровневых представлений групп эмбеддингов входной последовательности. За слоями CNN также следуют слой BiLSTM, слой внимания и два полносвязных слоя. Такая структура нейронной сети позволяет учитывать локальные характеристики фраз за счет CNN, а также семантику всего предложения за счет BiLSTM [14].

2.2. Классификаторы с использованием информации о тональности

Второй подход к обнаружению ироничных предложений использует информацию о тональности предложения в качестве вспомогательных данных. Ирония как литературный прием выражается в сокрытии или противопоставлении истинного смысла предложения его буквальному смыслу. Наиболее часто это проявляется тем, что вкладываемый в предложение эмоциональный окрас противоположен буквальному окрасу, т. е. текст, нацеленный на то, чтобы принизить или оскорбить, может выражаться фразами, имеющими вне контекста положительную тональность. Обнаружение такого несоответствия между буквальным и истинной тональностью предложения будет выступать в качестве показателя наличия ироничного посыла приложения.

С этой целью в корпус предложений, размеченных на предмет наличия иронии, дополнительно была добавлена разметка всех предложений по классам положительной, отрицательной или ней-

[1] https://ru.wiktionary.org

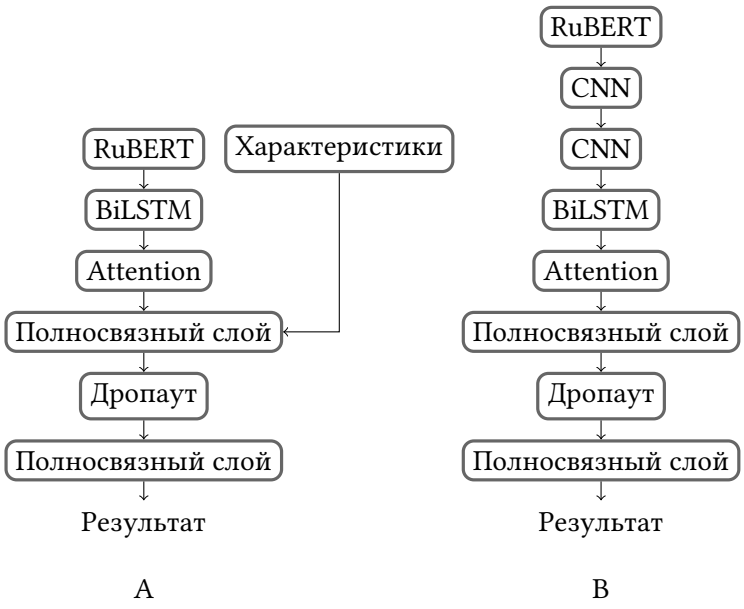

Fig. 2. A. RuBERT (LM) — BiLSTM — Att. + Fts. model. B. RuBERT (LM) — CNN — CNN — BiLSTM — Att. model

Рис. 2. A. Модель RuBERT (LM) — BiLSTM — Att. + Fts. B. Модель RuBERT (LM) — CNN — CNN — BiLSTM — Att.

тральной тональности. Разметка тональности подается на вход классификатору параллельно с разметкой по наличию или отсутствию иронии и используется на этапе обучения моделей.

Первая модель из данной группы представлена на рис. 3. Она имеет те же ключевые компоненты, что и ранее рассмотренные модели, но выполняет обработку двух типов входных данных параллельно. Как и ранее, с помощью RuBERT выполняется преобразование предложений в эмбеддинги, далее каждому эмбеддингу предложения сопоставляется соответствующая метка класса тональности и наличия иронии. Эти данные подаются на вход слоям BiLSTM: в один из них передаются эмбеддинги и метки наличия иронии, а во второй — те же эмбеддинги и метки тональности. Следом за каждой BiLSTM идет слой внимания; результаты, полученные на входе этого слоя конкатенируются и передаются в общую последовательность из двух полносвязных слоев. Эта цепочка слоев используется для обучения модели обнаружению иронии. Параллельно с этим из BiLSTM, в которую были переданы метки тональности, идет еще одна отдельная цепочка из полносвязных слоев. Ее результат используется только для обучения этой части модели классификации по тональности. На этапе оценки и разметки результаты классификации тональности игнорируются, так они не важны для данного исследования. В результате получается, что одна из BiLSTM обучается исключительно обнаружению иронии, в то время как вторая параллельно обучается классификации тональности, при этом передавая свои выходы в общую цепочку. Таким образом общая цепочка слоев работает как с информацией о наличии иронии, так и информацией о тональности предложений.

Вторая модель данной группы использует информацию о тональности предложений для дообучения модели RuBERT с целью получения эмбеддингов, изначально содержащих информацию об эмоциональном контексте употребляемых токенов (рис. 4A). Дообучение происходит с использованием обучающей выборки корпуса предложений текущего этапа кросс-валидации. После дообучения RuBERT также используется для создания контекстуализированных эмбеддингов, но их веса будут отличаться от весов эмбеддингов стандартного RuBERT в связи с проведенным дообучением. Полученные эмбеддинги вместе с информацией о наличии иронии передаются в модель из BiLSTM, слоя внимания и двух полносвязных слоев для обучения модели обнаружению иронии.

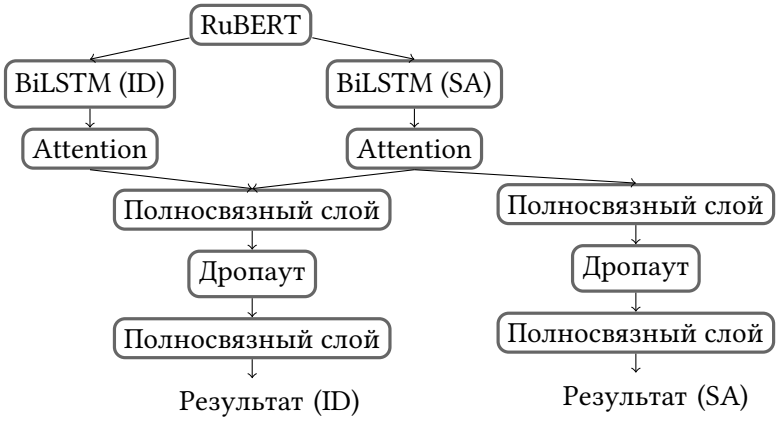

Fig. 3. RuBERT (LM) — BiLSTM (ID) — Att. + BiLSTM (SA) — Att. model

Рис. 3. Модель RuBERT (LM) — BiLSTM (ID) — Att. + BiLSTM (SA) — Att.

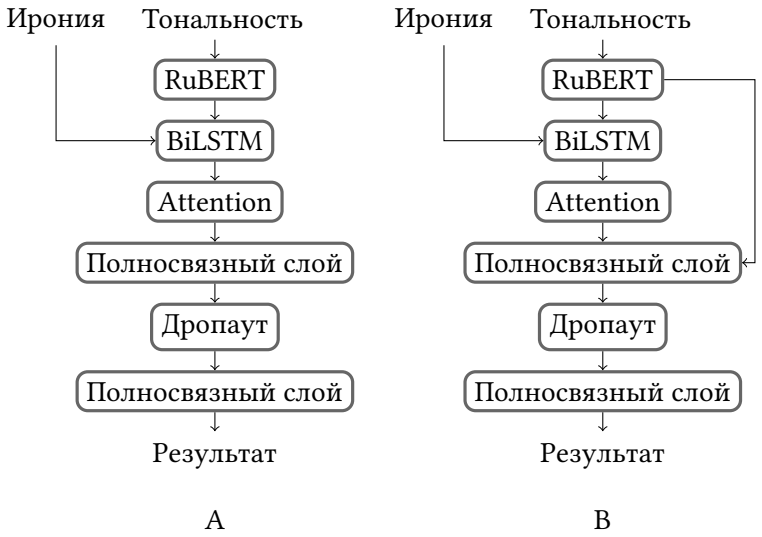

Fig. 4. RuBERT (SA) — BiLSTM — Att. model

Рис. 4. Модель RuBERT (SA) — BiLSTM — Att.

Таким образом, данная модель учитывает информацию о тональности предложения сразу, начиная с этапа создания эмбеддингов.

Следующая модель группы расширяет предыдущую, добавляя значения эмбеддингов токенов к значениям на выходе слоя внимания (рис. 4B). Идея данного подхода аналогична той, что описывалась для модели, приведенной на рис. 1C.

Еще одна модель данной группы использует ранее рассмотренную связку из эмбеддингов RuBERT, BiLSTM, слоя внимания и двух полносвязных слоев с тем различием, что вместо стандартной модели RuBERT для языкового моделирования применяется обученная на корпусе RuSentiment [15] модель RuBERT для классификации тональности. Использование такой модели также позволяет получать контекстуализированные эмбеддинги, учитывающие тональность.

2.3. Классификаторы с обучением эмбеддингов обнаружению иронии

К последней группе принадлежат классификаторы, в который RuBERT обучается задаче обнаружения иронии на корпусе ироничных предложений. Так как RuBERT изначально обучался задаче языкового моделирования на гораздо большем корпусе текстов из русскоязычной версии Интернет-ресурса Wikipedia и российских новостных изданий, то его дополнительное обучение задаче

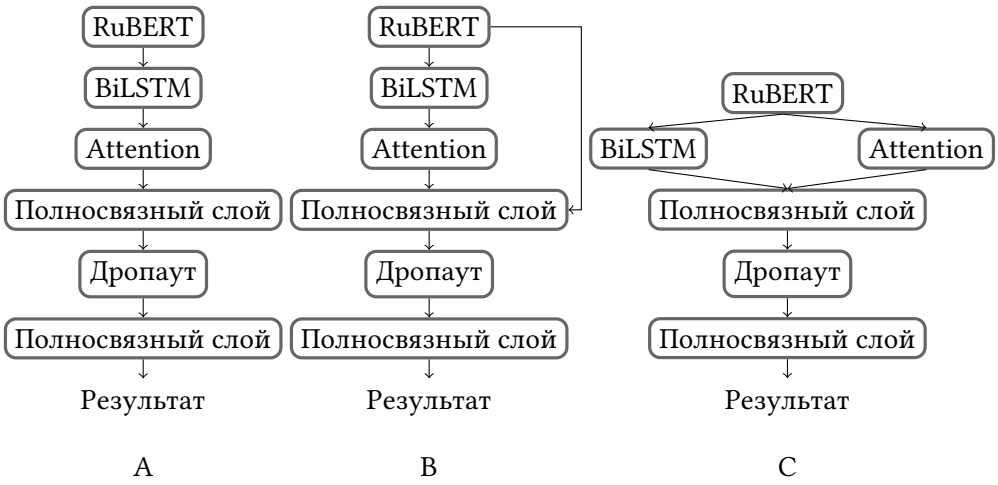

Fig. 5. RuBERT (ID)-based models featuring domain adaptation

Рис. 5. Группа моделей с адаптацией к предметной области RuBERT (ID)

обнаружения иронии на меньшем корпусе здесь является попыткой адаптации модели к заданной предметной области. В приведенных в данной группе классификаторах обучение RuBERT происходит совместно с обучением остальных слоев: каждое предложение проходит через цепочку из слоев RuBERT, BiLSTM и полносвязных слоев, в результате чего изменяются их веса, в том числе веса двенадцати внутренних слоев RuBERT.

Все модели данной группы приведены на рис. 5. Их структура аналогична ранее описываемым классификаторам с той лишь разницей, что веса внутренних слоев RuBERT не являются замороженными.

3. Корпуса текстов, используемые в экспериментах

Для проведения экспериментов в настоящей работе используется корпус русскоязычных текстов из новостных, аналитических статей и рецензий, отобранных в рамках исследования [1]. Корпус собирался из текстов, представленных на Интернет-ресурсе OpenCorpora[2]. Отобранные предложения проходили первоначальный этап разметки группой волонтеров, в ходе которого был составлен список предложений, в которых по оценке волонтеров выражается иронический или саркастический посыл. Далее выбранные волонтерами предложения прошли этап валидации экспертом, целью которого было уменьшение доли ошибок при разметке. Отобранные экспертом ироничные предложения были дополнены равным количеством предложений, не содержащих иронии. Так был составлен базовый корпус для экспериментов, содержащий 964 предложения каждого класса.

Для оценки влияния объема корпуса на качество классификации также рассматривался расширенный корпус, в который были добавлены предложения, приведенные в качестве примеров ироничного употребления слов из категории «Ироничные выражения/ru» ресурса Wiktionary. Для достижения баланса классов также было добавлено такое же количество предложений без иронии из корпуса OpenCorpora. Расширенный корпус содержит 1965 предложений каждого класса.

4. Результаты экспериментов

Результаты проведённых экспериментов приведены в таблице 1. Так как данная работа является продолжением работы [1], то полученные показатели будут сравниваться с показателями предыдущей работы. В качестве исходных показателей для сравнения берутся достигнутые в ней на базовом

[2] http://opencorpora.org

Table 1. Results of the experiments **Таблица 1.** Результаты экспериментов

Модель	Базовый корпус			Расшир. корпус		
	P	*R*	*F*	*P*	*R*	*F*
BERT	0.73	0.79	0.76	—	—	
BiLSTM	0.74	0.72	0.73	—	—	
BERT (LM) — BiLSTM	0.75	0.76	0.76	0.81	0.80	0.81
BERT (LM) — BiLSTM — Att.	0.85	0.75	0.80	0.82	0.82	0.82
BERT (LM) — BiLSTM — Att. + BERT (LM)	0.79	0.75	0.77	0.84	0.81	0.82
RoBERTa (LM) — BiLSTM — Att. + RoBERTa (LM)	—	—	—	0.84	0.84	0.84
BERT (SA) — BiLSTM — Att.	0.86	0.68	0.75	—	—	—
BERT (SA) — BiLSTM — Att. + BERT (SA)	0.87	0.72	0.79	—	—	—
BERT (fine-tuned, SA) — BiLSTM — Att. + BERT (fine-tuned, SA)	0.78	0.69	0.73	0.79	0.78	0.78
BERT (LM) — BiLSTM — Att. + BiLSTM — Att.	0.79	0.74	0.76	—	—	—
BERT (LM) — CNN — CNN — BiLSTM — Att.	0.82	0.74	0.78	0.82	0.82	0.82
BERT (ID) — BiLSTM — Att. + BERT (ID)	0.77	0.79	0.78	0.83	0.84	0.83
BERT (ID) — BiLSTM — Att.	0.73	0.79	0.76	0.74	0.84	0.78
BERT (ID) — Att. + BiLSTM	0.71	0.79	0.74	—	—	—
BERT (LM) — BiLSTM — Att. + Fts. (Punct.)	0.84	0.75	0.79	—	—	—
BERT (LM) — BiLSTM — Att. + Fts. (Punct., POS, Bigrams)	0.85	0.74	0.79	0.80	0.82	0.81
BERT (LM) — BiLSTM — Att. + Fts. (Punct., Irony Dictionary)	0.83	0.75	0.79	—	—	—

Обозначения: LM (Language Modelling) — моделирование языка; SA (Sentiment Analysis) — анализ тональности; ID (Irony Detection) — обнаружение иронии; Att. (Attention) — слой внимания; Fts. (Features) — использование дополнительных признаков: Punct. — пунктуации, POS — POS-тегов, Irony Dictionary — тонального словаря; *P* (Precision) — точность; *R* (Recall) — полнота; *F* (F-measure) — F-мера.

корпусе значения F-меры 0.76, полученное при дообучении модели RuBERT, и 0.73, продемонстрированное моделью BiLSTM с эмбеддингами spaCy[3].

Самая простая связка из RuBERT и BiLSTM на базовом корпусе показала результат, равный результату, полученному при использовании только модели RuBERT (F-мера — 0.76). Использование расширенного корпуса с этой связкой привело к одному из самых больших приростов F-меры — на 5 %. Обратная ситуация произошла со связкой, добавляющей слой внимания. На базовом корпусе она достигла F-меры 0.80, что стало лучшим результатом среди всех исследуемых моделей на базовом корпусе, но показала меньший прирост при использовании расширенного корпуса по сравнению с некоторыми другими классификаторами — 2 %. Дальнейшее усовершенствованием модели за счет добавления значений эмбеддингов к выходам слоя внимания не дало какого-либо прироста эффективности классификации ни на базовом корпусе (0.77), ни на расширенном (0.82).

Дальнейшего улучшения показателей классификации удалось добиться заменой языковой модели RuBERT на модель RuRoBERTa, которая обладает большим количеством внутренних слоев, параметров и размером эмбеддингов по сравнению с RuBERT. На расширенном корпусе данная модель достигла самого высокого результата среди всех проведенных экспериментов — 0.84. На базовом корпусе эксперимент с ней не проводился.

Среди моделей, использующих информацию о тональности, только одна показала результат близкий к удовлетворительному — связка из RuBERT, BiLSTM и слоя внимания с добавлением эмбеддингов. На базовом корпусе данная модель достигла F-меры 0.79, что на 1 % ниже лучшего результата на данном корпусе. На расширенном корпусе модель не оценивалась, так как дополнившие базовый корпус предложения не были размечены по классам тональности и ресурсов

[3] https://spacy.io/models/ru#ru_core_news_lg

на выполнение такой разметки у авторов работы не было. В целом низкие показатели классификаторов, использующих информацию о классах тональности предложений корпуса, объясняются значительным дисбалансом классов тональности. Более половины текстов имели нейтральную тональность, меньшая часть — отрицательную, и лишь очень малая часть предложений корпуса имела положительную тональность, что объясняется природой литературных оборотов иронии и сарказма.

Эксперименты с моделями, в которых RuBERT адаптировался под задачу обнаружения иронии, дали смешанные результаты. На базовом корпусе все они показали относительно низкие показатели в диапазоне 0.74–0.78, но на расширенном корпусе наиболее сложная модель данной группы, состоящая из RuBERT, BiLSTM и слоя внимания с добавлением эмбеддингов, дала второй по величине результат среди всех экспериментов — 0.83. Такой показатель может говорить о важности большого объема корпуса данных и наличия слоя внимания для адаптации модели BERT к предметной области.

Добавление в данные модели дополнительных характеристик (пунктуации, POS-тегов и т. д.) не оказало положительного влияния на показатели моделей: все классификаторы данной группы имеют почти одинаковые результаты с небольшими вариациями точности и полноты. Это может свидетельствовать о недостаточной существенности этих характеристик для решения задачи по сравнению с эмбеддингами.

Заключение

В данной работе описаны результаты исследования по разработке комплексных классификаторов для решения задачи обнаружения иронии на корпусах русскоязычных предложений. Был разработан и протестирован ряд моделей на двух корпусах текстов: базовом, используемом в предыдущей работе авторов, и расширенном. Результаты, продемонстрированные большинством разработанных классификаторов, превзошли результаты предыдущих экспериментов на базовом корпусе. Сравнение показателей моделей BiLSTM с BERT (LM) — BiLSTM и BERT (LM) — BiLSTM — Att. + BERT (LM) с RoBERTa (LM) — BiLSTM — Att. + RoBERTa (LM) явно показало, что увеличение размерности эмбеддингов, количества внутренних параметров в языковой модели и размера корпуса обучающих данных при переходе от эмбеддингов spaCy к BERT и от BERT к RoBERTa соответственно дает ощутимый прирост качества классификации модели без внесения каких-либо дополнительных модификаций в структуру. Таким образом, при наличии достаточных вычислительных мощностей можно добиться удовлетворительного качества классификации с относительно простым классификатором, если использовать большую языковую модель. Тем не менее, внедрение в модели средств выделения дополнительной информации из эмбеддингов все же показывает себя как наиболее эффективная доработка как в плане простоты реализации, так и в плане качества метрик.

Добавление в модель информации о тональности классифицируемых предложений не оказало желаемого эффекта на результаты экспериментов — все три модели данной группы показали ухудшение в метриках по сравнению с LM-моделями, что может быть вызвано дисбалансом классов тональности в размеченном корпусе. В целом составление корпуса, который был бы одновременно сбалансирован по классам тональности и классам наличия или отсутствия иронии, и обладал бы достаточным количеством предложений, является сложной задачей.

Добавление дополнительных характеристик, таких как наличие знаков пунктуации и POS-тегов, в модель также не привело к повышению метрик. В дальнейших экспериментах планируется использовать значительно более сложные характеристики (например, стилометрические), а также рассмотреть возможность учета языковых средств проявления иронии и сарказма, известных в классической лингвистике.

Для моделей, представляющих наибольший интерес, были проведены повторные эксперименты на расширенном корпусе. Во всех случаях F-мера увеличилась на 2–5 % при увеличении объема

обучающих данных, что характерно для нейронных сетей. Поскольку повышение качества классификации с объемом обучающей выборки все еще сильно сказывается на результатах классификации, составление подходящего корпуса продолжает оставаться одной из приоритетных задач.

Полученные в данной работе показатели являются лучшими на момент проведения исследования для задачи обнаружения иронии в русскоязычных текстах, а результаты наилучшего предложенного метода — RoBERTa (LM) — BiLSTM — Att. + RoBERTa (LM) — сравнимы со многими высокими показателями исследований для английского языка.

References

[1] M. Kosterin, I. Paramonov, and N. Lagutina, "Automatic irony and sarcasm detection in Russian sentences: Baseline methods", in *33rd Conference of Open Innovations Association FRUCT*, IEEE, 2023, pp. 148–154. DOI: 10.23919/FRUCT58615.2023.10142992.

[2] D. Šandor and M. B. Babac, "Sarcasm detection in online comments using machine learning", *Information Discovery and Delivery*, 2023. DOI: 10.1108/IDD-01-2023-0002.

[3] R. A. Potamias, G. Siolas, and A.-G. Stafylopatis, "A transformer-based approach to irony and sarcasm detection", *Neural Computing and Applications*, vol. 32, pp. 17 309–17 320, 2020. DOI: 10.1007/s00521-020-05102-3.

[4] C. Van Hee, E. Lefever, and V. Hoste, "Semeval-2018 task 3: Irony detection in English tweets", in *Proceedings of The 12th International Workshop on Semantic Evaluation*, 2018, pp. 39–50. DOI: 10.18653/v1/S18-1005.

[5] M. Khodak, N. Saunshi, and K. Vodrahalli, *A large self-annotated corpus for sarcasm*, 2017. arXiv: 1704.05579 [cs.CL].

[6] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang, "Sarcasm as contrast between a positive sentiment and negative situation", in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 704–714.

[7] S. Zhang, X. Zhang, J. Chan, and P. Rosso, "Irony detection via sentiment-based transfer learning", *Information Processing & Management*, vol. 56, no. 5, pp. 1633–1644, 2019. DOI: 10.1016/j.ipm.2019.04.006.

[8] D. Hazarika, S. Poria, S. Gorantla, E. Cambria, R. Zimmermann, and R. Mihalcea, *Cascade: Contextual sarcasm detection in online discussion forums*, 2018. arXiv: 1805.06413 [cs.CL].

[9] T. Zefirova and N. Loukachevitch, "Irony and sarcasm expression in Twitter", *EPiC Series in Language and Linguistics*, vol. 4, pp. 45–49, 2019. DOI: 10.29007/tpzw.

[10] A. Gurin and T. Zhukov, "Avtomaticheskoe opredelenie sarkazma v tekstakh na russkom yazyke", *Tsyfrovaya ekonomika*, vol. 1(22), pp. 44–53, 2023, in Russian.

[11] A. D. Yacoub, S. Slim, and A. Aboutabl, "A survey of sentiment analysis and sarcasm detection: Challenges, techniques, and trends", *International journal of electrical and computer engineering systems*, vol. 15, no. 1, pp. 69–78, 2024. DOI: 10.32985/ijeces.15.1.7.

[12] Y. Kuratov and M. Arkhipov, *Adaptation of deep bidirectional multilingual transformers for Russian language*, 2019. arXiv: 1905.07213 [cs.CL].

[13] D. Zmitrovich *et al.*, *A family of pretrained transformer language models for Russian*, 2023. arXiv: 2309.10931 [cs.CL].

[14] C. Zhou, C. Sun, Z. Liu, and F. Lau, *A C-LSTM neural network for text classification*, 2015. arXiv: 1511.08630 [cs.CL].

[15] A. Rogers, A. Romanov, A. Rumshisky, S. Volkova, M. Gronas, and A. Gribov, "RuSentiment: An enriched sentiment analysis dataset for social media in Russian", in *Proceedings of the 27th international conference on computational linguistics*, 2018, pp. 755–763.