

**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук

УТВЕРЖДАЮ  
Академический руководитель  
образовательной программы  
«Науки о данных»,

\_\_\_\_\_ С.О. Кузнецов  
« \_\_\_\_ » \_\_\_\_\_ 201\_ г.

**Выпускная квалификационная работа**

на тему

Глубокое обучение в аспектном анализе тональности текстов

---

тема на английском языке

Deep Learning for Aspect-Based Sentiment Analysis

---

по направлению подготовки 01.04.02 «Науки о данных»

Научный руководитель	Выполнил
к.ф.–м.н., доцент	студент группы ИССА
Должность, место работы	2 курса магистратуры
доцент	образовательной программы «Науки о данных»
ученая степень, ученое звание	Г.О. Соколов
Е.И. Большакова	И.О. Фамилия
И.О. Фамилия	Подпись, Дата
Оценка	
Подпись, Дата	

**Москва 2017**

## Содержание

1.	Введение .....	3
1.1.	Актуальность темы .....	3
1.2.	Постановка задачи .....	6
1.3.	Область применения методов глубокого обучения.....	8
1.4.	Потенциальные проблемы .....	10
1.5.	Цель работы.....	13
2.	Обзор методов решения задачи .....	14
2.1.	Извлечение аспектов .....	15
2.1.1.	Традиционные методы .....	15
2.1.2.	Машинное обучение .....	19
2.1.2.1.	Стандартные методы машинного обучения.....	20
2.1.2.2.	Глубокое обучение .....	23
2.2.	Группировка аспектов .....	25
2.3.	Определение тональности аспектов.....	27
2.3.1.	Традиционные методы .....	28
2.3.2.	Машинное обучение .....	30
2.3.2.1.	Стандартные методы машинного обучения.....	30
2.3.2.2.	Глубокое обучение .....	32
2.4.	Извлечение прочих сущностей текста .....	34
3.	Проведение и анализ экспериментов .....	36
3.1.	Описание данных .....	36
3.1.1.	Данные SentiRuEval .....	36
3.1.2.	Дополнительные данные .....	39
3.2.	Постановка задачи и метрики качества .....	40
3.3.	Базовое решение .....	41
3.4.	Модели для экспериментов .....	43
3.4.1.	Word2Vec .....	45
3.4.2.	Сверточные нейронные сети.....	46
3.4.3.	Рекуррентные нейронные сети .....	49
3.5.	Оценка результатов .....	54
3.5.1.	Выбор оптимальной модели .....	54
3.5.2.	Использование дополнительных признаков .....	60
3.6.	Рекомендации для дальнейших исследований .....	63
	Заключение.....	64
	Список литературы.....	65

# 1. Введение

## 1.1. Актуальность темы

Анализ тональности текстов, т.е. определение мнения автора по отношению к объекту [44], в последнее десятилетие является распространенной задачей автоматической обработки естественного языка. Актуальность этого направления во многом связана с развитием социальных сетей, онлайн-рекомендательных сервисов, содержащих большое количество мнений пользователей по разным вопросам, в частности, о товарах и услугах.

Практическая потребность в эффективном анализе большого количества текстов первоначально появилась со стороны онлайн-бизнеса с целью улучшения качества обслуживания клиентов. Высокий уровень конкуренции среди онлайн-ритейлеров, продающих однотипную продукцию, мотивирует компании искать дополнительные способы повышения лояльности покупателей. Менее активно аналогичными инструментами пользуются некоторые игроки ресторанного и банковского бизнеса. В названных предметных областях эффективность применения автоматического анализа мнений пользователей достигла наибольшего результата. Авторы прикладных исследований, которые предлагают новые методы анализа тональности текстов, в основном пользуются данными из этих предметных областей [9, 31].

Растущий спрос на подобные системы подтолкнул к повсеместному созданию стартапов и подразделений внутри крупных компаний, разрабатывающих свои технологии в данной области – Microsoft, Google, Hewlett-Packard, SAS, Сбербанк и т.д. С другой стороны, интерес к анализу тональности мнений также существует со стороны научных организаций и университетов, где часто делается акцент на анализе социальных взаимоотношений внутри общества.

В последние годы спрос на продуктовые решения по данному направлению создается даже государственными структурами [23]. Предварительная оценка рейтинга власти, результатов выборов – примеры применения современных методов анализа мнений граждан. К менее открытым для общественного внимания относятся проекты, связанные с угрозой государственной безопасности – борьба с экстремизмом и терроризмом.

Источниками информации для массового анализа мнений являются социальные медиа (обзоры, обсуждения на форумах, блоги, социальные сети, комментарии на сайтах), которые переживают взрывной рост в последнее десятилетие. Индивидуальные потребители и компании все активнее пользуются этим контентом для принятия решений. Сегодня пользователь, собирающийся приобрести товар или услугу, благодаря огромному количеству обсуждений в Интернете, не ограничен мнением только ближайшего круга людей – членов семьи, родственников, друзей. Компании, в свою очередь, меньше нуждаются в проведении опросов, голосований, создании фокус-групп для получения мнений клиентов.

Определение тональности высказываний на уровне отдельных аспектов – сравнительно более современная и недавняя тенденция. Так стандартными аспектами для сферы ресторанного обслуживания являются *качество кухни, обслуживание, стоимость, местоположение*. Аспектный анализ предоставляет более детальную информацию о мнениях пользователей и позволяет создавать еще более эффективные решения выявленных проблем, являясь в полном смысле автоматическим анализом. Например, в случае традиционной классификации отзывов на положительные и отрицательные, эксперту на следующем этапе приходится вручную обрабатывать негативные оценки и определять критические зоны недовольства клиентов. Аспектный анализ, автоматически локализует выявленные проблемы, однако является более сложным в реализации.

До недавнего времени большинство аналогичных или похожих задач решалось, как правило, с использованием традиционных инженерно-лингвистических методов, основанных на правилах и особенностях конкретного языка. В последние несколько лет в связи с преодолением проблемы недостатка вычислительных ресурсов резко возрос интерес к современным методам машинного обучения, которые в состоянии сократить время решения задачи и участие профессионалов-лингвистов при сопоставимом качестве результатов.

Новейший интерес к подходам машинного обучения неразрывно связан с глубоким обучением (или глубинным обучением, от англ. *deep learning*). Алгоритмами глубокого обучения принято считать класс моделей машинного обучения, которые производят ряд последовательных нелинейных преобразований с входными данными для решения общеизвестных задач классификации, регрессии и т.п. В качестве побочного продукта таких преобразований получается высокоуровневое представление исходных данных. На практике, моделями глубокого обучения называют нейронные сети с несколькими скрытыми слоями. Некоторые распространенные примеры – многослойный перцептрон (MLP), сверточные (CNN) и рекуррентные нейронные сети (RNN). В рамках данной работы, термины «глубокое обучение» и «нейронные сети» (или «нейросети») будут считаться взаимозаменяемыми понятиями.

Главными преимуществами методов глубокого обучения являются большая гибкость модели, т.е. меньшая зависимость от конкретного языка и предметной области, а так же способность более качественно учитывать контекст предложений благодаря архитектуре некоторых нейронных сетей. Одним из наиболее ярких недавних примеров использования глубокого обучения в IT-индустрии – начало использования Google Translate нейронных сетей для машинного перевода [28]. Многие исследования также показывают преимущество в качестве, которое достигается за счет использования подходов глубокого обучения в некоторых задачах.

Учитывая потенциальные преимущества глубокого обучения, необходимо также хорошо понимать область и границы применимости этих методов. Зачастую нейронные сети рассматриваются как панацея для универсального решения любой задачи. Однако, стоит понимать, что данные подходы, как и другие алгоритмы машинного обучения, имеют собственные ограничения.

## 1.2. Постановка задачи

Обычно, формализуя задачу аспектного анализа, говорят о поиске в тексте пятёрок вида  $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$  [23], где:

- $e_i$  – название объекта, для которого существует мнение пользователя,
- $a_{ij}$  – аспект какого-либо объекта,
- $s_{ijkl}$  – мнение некоторого пользователя в отношении аспекта  $a_{ij}$  в некоторый момент времени (может быть положительным/отрицательным/нейтральным либо в соответствии со шкалой, например от 1 до 5),
- $h_k$  – конкретный пользователь, обладающий мнением,
- $t_l$  – время создания мнения пользователем.

Данное определение является обобщением для широкого круга задач, связанных с анализом тональности текстов. Например, традиционная классификация отзывов на позитивные и негативные является аспектным анализом с единственным аспектом – *общее* отношение пользователя к объекту в целом. Таким образом, стандартная классификация мнений на 2 класса является частным случаем сформулированной выше задачи.

Но, строго говоря, данная формулировка не до конца учитывает все нюансы аспектного анализа. Например, аспектные термины могут обладать собственными признаками. Так, если объектом отзыва является мобильный телефон, пользователь может высоко оценить качество звука гарнитуры, но высказаться негативно об её цене. В данной ситуации цена представляет собой аспект гарнитуры, которая, в свою очередь, является аспектом объекта «телефон». В иных случаях, аспект обладает сразу несколькими

тональностями, в зависимости от условий. Шумное кафе с живой музыкой может идеально подходить для компании молодых людей, но быть неприемлемым местом для деловой встречи.

Таким образом, первоначальная формулировка задачи не является исчерпывающей для аспектного анализа, однако применима относительно большинства задач и позволяет эффективно структурировать текстовую информацию. Именно такой формулировки придерживается большинство исследователей в своих работах.

Стоит сказать, что названную сущность  $h_k$  (информация о пользователе) стоит понимать в широком смысле, а не просто как имя автора. На практике важно знать всю доступную информацию об авторе – пол, возраст, география, активность и т.п.

В прикладных задачах подобного рода также бывает важно оценить общее отношение автора к объекту. Считается, что это мнение явно отражено в тексте и выделяется как один из аспектов помимо прочих. Однако, это не всегда так, поэтому в данной работе данная подзадача выделена в отдельный этап.

Решение формализованной выше задачи можно представить в виде последовательности шагов, изображенных на рисунке 1:



*Рисунок 1. Этапы аспектного анализа.*

Важной особенностью работы с неструктурированными текстами, написанными на естественном языке, является тот факт, что для обозначения одинаковых сущностей могут использоваться различные слова или выражения. Например, синонимы и местоимения указывают на один и тот же предмет, что легко распознаётся человеком, но представляет сложность для компьютерного алгоритма. В отзыве о новом мобильном телефоне объект может представляться словарными единицами «телефон», «смартфон» или даже «он», хотя речь все время идет об одном и том же предмете. В результате, после выделения набора всех возможных объектов, аспектов, субъектов, необходимо провести их группировку, что отражено в названиях первых двух этапов на рисунке выше.

Формулировка задачи является обобщенной, поэтому на практике некоторые шаги могут быть пропущены. Например, не всегда требуется знать время создания текста или информацию об авторе. При этом, данная информация часто может быть легко извлечена, так как тексты на форумах, в социальных сетях и т.п. содержат её в явном виде. В случае необходимости, эти сведения позволяют построить динамику тональности во времени или получить желаемый срез, например по географическому, половому или возрастному признаку.

В реальности тональность текста в целом или по отношению к аспектам не всегда обладает полярностью, например в высказываниях о фактических свойствах объекта/аспекта: *«Корпус телефона темно-серого цвета.»*. Другой частый случай – отсутствие упоминания аспекта в тексте.

### **1.3. Область применения методов глубокого обучения**

Использование подходов глубокого обучения в задаче аспектного анализа начинается, как правило, с поиска векторных представлений слов документа. Вектора слов с похожим контекстом расположены рядом с друг



другом (например, с точки зрения косинусного расстояния). Для векторизации слов часто используются такие известные подходы, как Word2Vec [29] и GloVe. Первый из них как раз представляет собой нейронную сеть с одним слоем, в котором каждому слову из заранее заданного словаря ставится в соответствие вектор необходимой размерности. Стоит сказать, что из-за простоты архитектуры Word2Vec иногда не относят к классу методов глубокого обучения. С другой стороны, обучение даже такой несложной нейросети на огромных текстовых коллекциях было бы невозможно без некоторых технических приемов, поэтому логично называть Word2Vec глубоким обучением. В контексте общей задачи, векторизацию слов можно считать подготовительным этапом, так как, например, полученные векторные представления можно использовать как альтернативу «мешку слов» для обычной логистической регрессии.

Основной подзадачей аспектного анализа, где применение нейросетей может быть крайне полезным, принято считать извлечение аспектных терминов. В литературе обычно это рассматривается как частный случай задачи по разметке последовательности слов/токенов – *sequence labeling*. Например, другой похожий пример *sequence labeling* – это разметка частей речи в предложении. Возможность учитывать порядок следования слов в предложении интуитивно является заметным преимуществом перед стандартными методами машинного обучения. Самым распространенным альтернативным решением, не связанным с глубоким обучением, является использование условных случайных полей – *conditional random fields (CRF)*. Демонстрируя часто высокое качество на практике, этот метод требует создания большого количества признаков, что необходимо делать вручную и каждый раз заново для конкретного языка. Применение нейросетей призвано решить эту проблему.

Глубокое обучение, помимо поиска аспектов, также используется для определения тональности аспектов или текста в целом. Эта задача несколько чаще встречается в исследованиях по автоматической обработке

естественного языка, и у нейронных сетей здесь может быть много «конкурентов», основной среди которых – *SVM*. Однако, все та же способность задействовать последовательности слов может быть преимуществом для методов глубокого обучения.

## 1.4. Потенциальные проблемы

Как уже было отмечено, анализ тональности текстов в разрезе аспектов – перспективная, но трудоемкая задача. Трудности, с которыми можно столкнуться, условно разделены на 3 группы:

1. Определение взаимоотношений между объектами, их частями и аспектами.
2. Корректность определения тональности.
3. Проблема данных и область применимости финальной модели.

Первая из названных проблем находится в плоскости семантического анализа и характерна для многих прикладных задач автоматизированной обработки текстов на естественных языках. В качестве примера можно привести предложение *«качество изображения камеры Canon оставляет желать лучшего»*. Объект в данном предложении – *камера Canon*, принадлежащий ему аспект – *качество съёмки/изображения*. Компьютерная программа, вероятно, определит фразу *«качество изображения»* как самостоятельный объект.

Исследователи обычно выделяют явные и неявные аспектные термины [23]. В первом случае, аспектами обычно являются существительные, которые указываются явно в тексте – *«новый айфон будет продаваться по очень высокой цене»* – аспект *цена* задан явно. С другой стороны, неявные признаки часто обозначаются прилагательными и наречиями – *«новый айфон будет стоить дорого»*. Аналогично, для человека не составит труда понять, что речь идет о цене телефона, но алгоритм может легко допустить ошибку, считая, что аспектом является общее мнение о продукте в негативной интонации.

Вторая проблема, связанная со сложностью определения тональной направленности, иногда происходит из-за выхода конкретной задачи за рамки обозначенной формулировки. Ранее уже было отмечено, что некоторые суждения бывают справедливы лишь при определенных условиях – *«новый айфон станет хорошим подарком для молодых людей»*. Высказывание справедливо и носит положительный оттенок не в целом, а лишь в конкретной ситуации, когда речь идет о молодежи. Отдельной проблемой может стать наличие сравнений в тексте – *«андроид работает намного быстрее»*. Алгоритм, в условиях строгой формулировки задачи, в лучшем случае проигнорирует данную фразу, в худшем – извлечет новый аспект айфона – *андроид*, и присвоит ему положительную тональность.

Часто мнение автора не сопровождается словами с эмоциональной окраской (*любить, ненавидеть и т.п.*), а изложено в форме свершившегося факта – *«наушники сломались на второй день после приобретения»*. С другой стороны, фактические высказывания не всегда содержат оценочные суждения – *«я приобрёл айфон позавчера»*. Важно правильно различать части текста, не содержащие суждения, то есть имеющие нейтральную тональность. Многие авторы также выделяют в отдельную группу категорию отзывов, связанных с потреблением ресурсов объектами. Это крайне актуально, в частности, для электронных устройств. Простой пример – *«новый айфон быстро расходует батарею»*.

Дополнительные трудности с классификацией тональности суждений могут возникнуть в ситуации, когда текст содержит сленг, иронию или сарказм. Последнее особенно характерно для коротких сообщений, написанных в Твиттере. Существующие подходы и алгоритмы пока показывают довольно низкую точность классификации при высоком содержании саркастических оборотов в текстах [23].

Под каждым текстом отзыва обычно требуется провести финальную черту, построив интегральную оценку высказывания. Например, большинство пользователей могут негативно высказываться о конкретном

аспекте. Тем не менее, общая тональность по отношению к объекту вполне может оставаться позитивной. Самым простым случаем определения интегральной оценки является ситуацию, когда тексте содержит суждение об объекте в целом. Однако, часть отзывов содержит отношение автора лишь к отдельным аспектам. В этом случае, определение интегральной оценки несколько усложняется [39].

Оценка качества модели – важный момент исследования. Некоторые тексты крайне сложно классифицировать даже человеку. Лучшая из предложенных практик – сравнивать качество алгоритма с усредненными оценками группы экспертов. Такой подход исключает смещённость результатов полученной модели из-за субъективности экспертного мнения.

В случае использования подходов, связанных с машинным обучением, отсутствие размеченных данных делает невозможным процесс тренировки модели. Если применять нейросети, проблема только усугубляется, так как такие алгоритмы более параметризованные и сложные и, следовательно, требуется еще больше данных для их качественного обучения.

Другая сложность заключается в ограниченности применения приложения в рамках одной области. Например, компания, реализующая на рынке широкий ассортимент товаров, не сможет одинаково использовать выработанный алгоритм для оценки мнений по всем своим продуктам. Сегодня, особенно хорошо проработанными считаются предметные области, связанные с электроникой, ресторанным и банковским бизнесом. Переход из одной области применения в другую или смена целевого языка обычно сопровождается созданием новых моделей с нуля. Использование методов глубокого обучения может помочь сделать алгоритмы более гибкими.

## 1.5. Цель работы

Данная работа полностью посвящена задаче аспектного анализа тональности текстов. Одна из целей данного исследования – показать актуальность и перспективность темы в свете развития информационных технологий и, в особенности, современных подходов и методов машинного обучения, а также провести обзор основных решений, предложенных другими авторами. Спектр представленных подходов очень широк и варьируется от использования простых лингвистических правил и шаблонов до применения глубоких нейронных сетей. Крайне важно понимать область применения, преимущества и ограничения, связанные с тем или иным методом.

Несмотря на стремительный рост популярности методов глубокого обучения в последние несколько лет, основные исследования авторов посвящены задачам распознавания образов, генерации изображений, обучению с подкреплением. Из-за относительно небольшого объема работ по аспектному анализу сложно объективно оценить, насколько нейросети могут помочь в решении данной задачи. Поэтому важно уделять больше внимания этой тематике, чтобы лучше понимать область и границы применимости методов глубокого обучения. В оптимистичном сценарии, такого рода исследования способны натолкнуть авторов на новые мысли по совершенствованию и созданию новых архитектур нейросетей для автоматической обработки естественного языка.

В практической части данной работы были проведены эксперименты с использованием различных архитектур нейронных сетей и некоторыми дополнительными признаками – части речи, знаки эмоций и пунктуация. Для обучения модели и оценки качества использованы размеченный набор отзывов конкурса *SentiRuEval-2015* [39] на русском языке. В завершение, предложен набор рекомендаций по дальнейшему улучшению результатов.

## 2. Обзор методов решения задачи

Существующие подходы, предложенные авторами ранее, наиболее удобно рассматривать в контексте подзадач, которые возникают на разных этапах аспектного анализа (см. рис. 1). Как уже было отмечено ранее, основными этапами являются извлечение и определение тональности аспектных терминов. Эти подзадачи будут рассмотрены более подробно с выделением в отдельную группу исследований, которые опираются на методы машинного обучения.

В рамках подходов машинного обучения обычно принято выделять группы алгоритмов, основанных на обучении с или без учителя. Строго говоря, использование нейронных сетей для поставленной задачи является частным случаем машинного обучения с учителем, когда по обучающей размеченной выборке тренируется модель и минимизируется выбранный функционал ошибок. Однако, тема данной работы непосредственно связана именно с глубоким обучением, поэтому данные методы несколько искусственно будут выделены в отдельную группу для повествования. Это не совсем соответствует устоявшейся терминологии, но позволяет отделить выбранную область интересов и сфокусироваться на ней. Остальные методы машинного обучения (с учителем и без) будут называться далее *стандартными методами машинного обучения*, так как на практике они используются более давно и, как правило, устроены несколько проще нейронных сетей. Подходы, основанные на лингвистических правилах, поиске языковых связей в тексте, использовании словарей и даже частотном анализе для простоты понимания будут называться *традиционными методами*.

## 2.1. Извлечение аспектов

Обзор данного этапа в основном касается явных аспектов, которые напрямую указываются в тексте – *«стоимость обучения в Высшей Школе Экономике чрезвычайно высока»*. Для сравнения, в случае неявного аспекта: *«ВШЭ – дорогое учебное заведение»*. Как было отмечено выше, извлечение аспектных терминов будет рассмотрено с точки зрения двух основных подходов:

1. Традиционные методы.
2. Машинное обучение.

В рамках второго подхода отдельно выделена группа исследований с применением глубокого обучения.

### 2.1.1. Традиционные методы

Большинство классических подходов, которые достаточно давно используются для извлечения аспектных терминов, основаны на частотном анализе либо на поиске связей между суждением и соответствующим объектом.

В работе [9] авторы выделяют существительные и группы существительных (обычно не более 3-4 слов) используя принадлежность слов к частям речи. Затем тривиально подсчитывается частота употребления выделенных конструкций и выбираются наиболее представительные из них. Порог отсека является параметром алгоритма и подбирается экспериментально. Относительно простой подход, тем не менее работает на практике, так как при агрегации большого количества отзывов, несмотря на разнообразие словоупотреблений внутри каждого из них, выделяются общие для всех слова или группы слов. Данные конструкции, в большинстве случаев, совпадают с реальными аспектами объекта.

Предложенный подход был модифицирован в работе [31]. Авторы попытались отфильтровать лишние слова, которые не могут являться аспектными терминами, используя значение показателя *PMI*:

$$PMI(a, d) = \frac{hits(a \wedge d)}{hits(a)hits(d)},$$

где *a* – некоторый аспект-кандидат, *d* – конструкция-связка для поиска меронимов объектов, например «*айфон в комплекте с*», «*айфон имеет*», «*айфон обладает*» для объекта «*айфон*».

Идея используется во многих других задачах автоматической обработки естественного языка и заключается в определении насколько часто встречаются два слова вместе, чем по отдельности. Для подсчета значения частоты словоупотреблений создавались соответствующие поисковые запросы в Интернете. Дополнительно, авторы пользовались ресурсом WordNet для поиска отношений часть-целое. Подобный подход является затратным с точки зрения времени, поскольку происходит интенсивное обращение к интернет-поиску. Кроме того, метод считается малоприменимым в отношении к фильмам, ресторанам, программному обеспечению.

Другой подход, основанный на частоте употреблений был продемонстрирован в работе [20]. Авторы использовали популярный метод TF-IDF применительно к потенциальным аспектам на уровне абзаца и всего документа.



В [53] был апробирован метод извлечения аспектов, состоящих из нескольких слов на основе метрики C-value, описанной ранее в другой работе [5]. В отличие от стандартной частоты, метрика учитывает длину аспекта-кандидата и другие конструкции, которые также включают в себя анализируемый аспект. C-value снижает вес данного слова или словосочетания, если оно входит в частотное словосочетание большей длины. Тем самым предполагается, что более длинное словосочетание может рассматриваться как кандидат на аспект, а текущее представляет его фрагмент.

Long и др. [25] попытались извлечь аспекты, используя частоту и информационное расстояние. Выбирались наиболее близкие к данному кандидату слова с точки зрения информации, которые затем использовались для поиска определенных текстов, обсуждающих в наибольшей степени данный аспект.

Если в качестве аспектных терминов извлекаются не только отдельные существительные, но и именные группы, то необходимо использовать дополнительные признаки для более точного определения длины именной группы. Чаще всего используются так называемые контекстные признаки, которые оценивают частоту встречаемости словосочетания с частотой контекста. Такие признаки позволяют определить границы именной группы. Например, в [2] используется мера FLR:

$$FLR(a) = f(a) \cdot LR(a), \quad LR(a) = \sqrt{l(a) \cdot r(a)},$$

где  $f(a)$  – частота аспектного термина,  $l(a)$  – количество разных слов, находящихся слева от  $a$ ,  $r(a)$  – количество разных слов справа от  $a$ .

Далее отбираются группы существительного с данной мерой, большей, чем в среднем для словосочетаний. Таким образом, данная мера в первую очередь отбирает группы существительного, которые имеют большое разнообразие слов на своих границах, что показывает, что анализируемый термин  $a$  не является фрагментом более длинного словосочетания.

Если текст содержит суждения, то, логично предположить, что они относятся к какому-то объекту. Пользуясь этим простым правилом, Ну и др. [9] выявляли редко встречающиеся аспекты. Идея заключается в следующем – если предложение не содержит частотные аспекты, производится поиск существительных или именных групп в непосредственной близости к словам с эмоциональной окраской. Этот довольно тривиальный метод показывает неплохие результаты на практике. Дополнительно, он может применяться для выделения только важных аспектов объекта, так как тональность в тексте обычно присутствует только в отношении действительно значимых аспектов.

В работе [54] для извлечения отношений между аспектными терминами и оценочными словами используется синтаксический анализатор. Отношения между аспектом и оценочным словом извлекаются на основе заданных путей синтаксической зависимости. Так, например, в предложении «*This movie is not a masterpiece*» слова *movie* и *masterpiece* будут размечены соответственно аспектом и оценочным словом, поскольку между ними существует путь в синтаксическом дереве «*NN – nsubj – VB – dobj – NN*».

Для извлечения аспектных терминов с учетом их отношений с оценочными словами часто используются итеративные методы (*bootstrapping*). В качестве начального множества могут использоваться частотные именные группы, которые предполагаются аспектами либо задаются вручную.

В известной работе [9] начальное множество аспектных терминов (частотные слова и именные группы) используется для выявления ассоциативных правил, т. е. шаблонов, посредством которых аспекты обычно связаны с оценочными словами. После получения таких правил извлекаются менее частотные аспектные термины, т.е. те именные группы, которые появлялись именно в таких шаблонах с оценочными словами.

В исследовании [33] рассматривается подход двойного распространения (double propagation) к извлечению аспектных терминов и расширению словаря оценочных слов. В качестве исходного множества задается небольшой словарь оценочных слов, а также синтаксические шаблоны, в которые обычно входят оценочные слова и аспектные термины. В итоге вхождение известного оценочного слова в такой шаблон помогает извлекать аспект, а известный аспект, входящий в шаблон, помогает извлекать оценочное слово.

Для очистки полученного множества аспектов применяется ряд правил. Например, предполагается, что в одном фрагменте предложения без запятых содержится только один аспектный термин, а другой кандидат должен быть удален, удаляется менее частотный в коллекции.

В работе [50] для оценки значимости аспектных терминов вводятся еще два фактора. Первый фактор рассматривает, насколько разнообразны оценочные слова, применяемые к аспекту-кандидату, – разнообразие обычно свидетельствует о значимости аспектного термина. Во-вторых, в коллекции ищется подтверждение связи аспектного термина с сущностью посредством заданных шаблонов. Например, в области автомобилей можно найти такие фразы, как *«двигатель автомобиля»*, *«автомобиль имеет большой двигатель»*, которые свидетельствуют об отношении часть–целое между словами *двигатель* и *автомобиль*. Если слово одновременно встречается и с оценочным словом, и в отношениях с заданной сущностью, то это дает этому аспекту-кандидату сразу большой вес: например, *«в матрасе имелась большая дыра»*.

### **2.1.2. Машинное обучение**

Работы, перечисленные в первом пункте данной главы, исследуют относительно новые, но достаточно устоявшиеся подходы машинного обучения. Во второй части кратко изложены публикации, затрагивающие непосредственно модели глубокого обучения.

### 2.1.2.1. Стандартные методы машинного обучения

Этот раздел посвящен сразу двум основным парадигмам машинного обучения – с учителем и без учителя. Первая группа методов показывает на практике более качественные результаты и в первую очередь речь пойдет об этих подходах.

Извлечение аспектов может рассматриваться как частный случай задачи извлечения информации. Значительное количество алгоритмов было предложено в последние годы для решения последней проблемы. Подавляющее большинство методов основаны на последовательной разметке – *sequence labeling*. Такие модели нуждаются в первоначальных размеченных данных для обучения, то есть в ручном выделении аспектов и не-аспектов в корпусе. Самыми эффективными инструментами на сегодня являются скрытые Марковские модели (HMM) и условные случайные поля (CRF). Jakob и др. [12] обучили CRF на предложениях, взятых из обзоров различных областей. В качестве признаков были использованы токены, части речи, синтаксические зависимости, расстояния между словами и т.д.

В качестве признаков модели CRF авторы работы [11] использовали текущий токен и его часть речи, токены и их части речи внутри скользящего окна выбранной ширины, частоту токенов в обучающей выборке, наличие токена в специально сформированном словаре для заданной тематики.

Другие методы машинного обучения также успешно применяются в рамках данной задачи. Алгоритм, предложенный [18] выделяет сначала список пар аспект-оценка, используя деревья зависимостей, а затем решает традиционную задачу классификации с применением основанных на деревьях подходов. Затем из пар, получивших наивысшие вероятности, извлекаются финальные аспекты. Yu и др. [47] использовали одноклассовый метод SVM на данных, в которых были размечены только аспекты (без выделения не-аспектов).

Участники конкурса SentiRuEval [27, 54] также применяли SVM с L2 регуляризацией для выделения слов 3 типов – начало, конец аспектного термина или слово, не относящееся к аспектам. Для этого авторы использовали 3 вида признаков – локальные внутри предложения, глобальные, которые относятся ко всему документу в целом, и, наконец, признаки, которые связаны с внешними по отношению к тексту ресурсами.

В работе [19] для извлечения аспектных терминов помимо частотности аспектов-кандидатов в отзывах используется сопоставление кандидатов с заголовками словарных статей в Википедии, семантическая близость кандидатов, рассчитанная на основе совокупностей ссылок соответствующих статей Википедии (в итоге 2 признака), а также ассоциирование кандидата в аспекты с именем сущности при поиске в интернете.

Другая часто используемая парадигма – машинное обучение без учителя. Она позволяет избежать проблем, связанных с отсутствием обучающей выборки, что крайне актуально для задачи аспектного анализа. Помимо этого, методы обучения без учителя не используют большое количество признаков, которые необходимо генерировать для обучения с учителем. Очевидно, что обратная сторона названных преимуществ – худшее качество работы алгоритмов.

В последние годы были разработаны статистические тематические модели, которые используются для обнаружения тем внутри большой коллекции документов. Данный подход предполагает, что каждый документ состоит из набора тем, а каждая тема представляет собой вероятностное распределение среди слов. Данные модели относятся к классу генеративных моделей, которые основываются на некотором вероятностном механизме формирования текстов.

В первую очередь выделяют 2 основные модели – вероятностный латентно-семантический анализ (pLSA) и латентное размещение Дирихле (LDA). Несмотря на то, что они в основном применяются в задачах извлечения информации из текстов, могут быть легко адаптированы для работы с другими типами данных. Для аспектного анализа такие модели должны уметь выделять и различать одновременно аспекты и слова тональности.

Одна из известных модификаций базовой модели LDA для извлечения аспектных терминов описана в работе [41], где показано, что применение базовой модели LDA, которая строится на информации о взаимной встречаемости слов в одних и тех же текстах, не является эффективной для извлечения аспектов, поскольку во множестве разных отзывов может содержаться один и тот же набор аспектов. Авторы работы применяют глобальную модель для извлечения именованных сущностей, а для извлечения аспектных терминов используют скользящее окно из слов или предложений (например, 3 предложения). Собственно, встречаемость слов в таких фрагментах используется для выявления аспектов, при этом они не различают аспектные термины и оценочные слова.

В работе [51] предложена гибридная модель MaxEnt-LDA (комбинация моделей Maximum Entropy и LDA), в которой производится совместное извлечение аспектных и оценочных слов на основе синтаксических признаков, помогающих разделить аспектные и оценочные слова. Метод Maximum Entropy используется для подбора параметров на размеченных данных.

В [23] указываются следующие проблемы применения тематических моделей для извлечения и группирования аспектных терминов:

- требуются большие объемы данных и тщательная настройка параметров моделей для получения достаточно качественных результатов,

- методы основаны на семплировании Гиббса и поэтому каждый раз дают несколько иной результат,
- тематические модели не показывают ощутимо лучший результат, чем более простые методы, например, основанные на частотном анализе.

#### 2.1.2.2. Глубокое обучение

В практике применения глубокого обучения упомянутую ранее задачу *sequence labeling* обычно решают с помощью нейронных сетей, работающих по принципу *many-to-many*. То есть на вход нейросети поступает последовательность слов, и после каждого нового слова нейросеть делает предсказание, например, является ли данное слово аспектным термином. Таким образом на выходе из модели получается последовательность ответов, равная по длине входящей исходной последовательности слов/токенов в предложении.

В статье [40] были использованы различные архитектуры рекуррентных нейронных сетей (RNN) для извлечения аспектных терминов для отзывов на русском и английском языках для конкурсов SentiRuEval-2014 и SentiRuEval-2015. Выбранная 2-слойная LSTM-нейронная сеть с обратными связями (bidirectional 2-layer LSTM) показала заметно лучший результат, чем у CRF и по качеству оказалась в числе победителей конкурса. Что важно, было отмечено преимущество рекуррентных нейронных сетей для быстрого прототипирования с использованием других языков без необходимости генерации каких-либо новых признаков.

Авторы [32] применили сверточную нейронную сеть (CNN) с 2 слоями свертки, 2 слоями пулинга и полносвязным слоем. Эксперименты на различных наборах данных показали, что дополнительное использование лингвистических правил и информации о частях речи может значительно улучшить качество модели. Кроме того, применение векторного представления слов, обученного на специализированном корпусе даже меньшего размера, чем корпус общий направленности, также существенно усиливает модель. В частности был использован корпус из около 35 млн отзывов (4.7 млрд слов) на Amazon вместо огромного корпуса Google News (около 100 млрд слов).

Xu L. и др. [45] аналогично обучили CNN, используя данные отзывов Yelp. Для извлечения аспектов авторы решали задачу классификации. Порог принятия решения при этом был параметризован и оптимизировался наряду с остальными гипер-параметрами модели.

В [21] исследователи использовали деревья синтаксического разбора предложений совместно с RNN для получения векторных представлений слов, которые учитывают зависимости на уровне фраз/предложения. Затем данные векторы были использованы для оптимизации функции потерь в задаче поиска пар аспект-тональность. Примечательно, что задачи извлечения аспектных терминов и определения их тональности решались одновременно, что позволило улучшить результаты по сравнению с независимым подходом в два этапа.

Jebbara S. и Cimiano P. [13] применили архитектуру GRU (Gated Recurrent Units) с обратными связями. Авторы попытались учесть дополнительную информацию из WordNet в векторах слов, полученных на большом корпусе отзывов (около 83 млн текстов). Для этого решалась задача оптимизации – связанные в графе WordNet слова должны иметь близкие векторные значения, при этом не отклоняясь от исходных векторов. Данный подход, однако, никак не помог улучшить качество поиска аспектов.



Авторы [43] применили достаточно простую архитектуру полносвязной нейронной сети, обученную на текстах ресторанной тематики. Для каждого слова в предложении была предсказана вероятность его принадлежности к той или иной заранее заданной аспектной группе. Эксперименты показали, что даже прямое использование этой модели без адаптации для гостиничных отзывов показывает отличные результаты.

Примечательным является использование Toh Z. и Su J. [42] глубокого обучения (RNN с обратными связями) для генерации дополнительных признаков, необходимых CRF. Для обучения RNN помимо стандартных векторов слов использовались также кластеры, полученные из этих векторов с применением алгоритма K-means. Получившиеся кластеры отражают высокоуровневые понятия языка.

В работе [52] предложен альтернативный подход поиска векторного представления слов, которое учитывает различные связи в предложении. По словам авторов, их метод превосходит Word2Vec и Glove по качеству. Было задействовано несколько MLP архитектур для создания признакового пространства с последующим обучением логистической регрессии.

## **2.2. Группировка аспектов**

На практике, для решения задачи аспектного анализа заранее задан набор аспектных групп – укрупненных характеристик объекта. Такими характеристиками для ресторанной тематики могут быть: *кухня, интерьер, обслуживание, местоположение*. Аспектные группы сами по себе состоят из большого числа достаточно разнообразных аспектных терминов. Так, например, группа «кухня» включает в себя названия блюд, продуктов питания и напитков, которые предлагаются в ресторане.

Использование общезначимых словарей синонимов и тезаурусов для этой цели имеет ограниченное применение, поскольку такие группировки аспектных терминов существенно зависят от предметной области. Кроме того, часто аспектные термины выражаются словосочетаниями, которые обычно не описываются в словарях.

В работе [51] показано, что автоматизация группировки аспектов является критической для многих приложений анализа тональности отзывов.

В статьях [48, 49] предложен алгоритм частичного обучения, который разбивает аспектные термины на predetermined категории аспектов. При этом предполагается, что сами по себе аспектные термины уже выделены каким-то методом. Сначала авторы вручную относят небольшое количество аспектных терминов к категориям. Затем применяют Expectation Maximization (EM) алгоритм для работы с размеченными и неразмеченными примерами. Кластеризация проводится на базе сходства контекстов упоминания аспектных в окне 15 слов влево и вправо. Если в окне встречается другой аспектный термин, то он не включается в окно. Также исключаются стоп-слова.

В методе также применяются два вида дополнительной информации для лучшей инициализации EM-алгоритма: аспектные термины в виде именных групп, имеющие общие слова, обычно относятся к одной категории аспектов (*battery life* и *battery power*), и аспектные термины, являющиеся синонимами в словаре, также чаще всего будут принадлежать одной группе. Эти две эвристики позволяют EM-алгоритму достигать лучших результатов.

В работе [46] ставится задача выстроить иерархическую классификацию аспектных терминов, подобно экспертной классификации. Иерархия аспектов строится на основе нескольких признаков сходства:

- контекстный признак: два слова влево и вправо,
- признак совместной встречаемости аспектных терминов, вычисляемый на основе меры взаимной информации PMI,

- длина синтаксического пути между аспектными терминами в предложении, а также синтаксические роли в предложениях (подлежащее, объект, модификатор и т. п.),
- лексические признаки, включая извлеченное из интернета определение аспектного термина.

Авторы в статье [52] для определения аспектной группы из заранее заданного списка решают задачу классификации логистической регрессией. Для этого они используют набор признаков, полученный с помощью двух типов обученных нейронных сетей с одним скрытым полносвязным слоем. Первый тип предсказывает вероятности каждого из пяти аспектных классов одновременно, другой тип – вероятность бинарного отклика для каждого из классов в отдельности. Таким образом, полученные 6 наборов весов со скрытых слоев конкатенируются и используются логистической регрессией.

## 2.3. Определение тональности аспектов

После извлечения из текста аспектных терминов им в соответствие ставится одна из тональностей. На практике отсутствует общеиспользуемый набор классов тональности. В конкурсах SemEval и SentiRuEval используется шкала из 3 значений – отрицательная, нейтральная либо положительная оценка автора. В некоторых исследованиях применяется бинарный отклик. Лишь немногие работы используют сложные шкалы, например оценку от 1 до 5. Это связано с тем, что значительно возрастает степень неопределенности, например, в связи с тем, что автор текста сам не очень хорошо понимает разницу между оценкой 2 и 3. Поэтому, качество модели может значительно ухудшиться в этом случае.

Как и в разделе, посвященном поиску аспектов, далее будут рассмотрены традиционные подходы и методы машинного обучения, наиболее часто встречающиеся в исследованиях. Под первыми обычно подразумевают инженерно-лингвистические модели.

### 2.3.1. Традиционные методы

В инженерно-лингвистических методах предполагается, что на момент классификации известны:

- названия сущностей, их аспектов;
- словарь оценочных слов и выражений, а также правила их преобразования в зависимости от контекста и правила суммирования. Обработка идет обычно по предложениям и включает в себя несколько этапов [23].

Сначала производится проставление в предложении известных аспектных терминов и оценочных слов; оценочные слова имеют проставленную в словаре оценку тональности. К оценочным словам применяются операторы, которые могут менять тональность оценочного слова на противоположную.

Далее необходимо учесть структуру предложения для возможной модификации базовых оценок. В частности, в работе [30] указывается на важность обработки союзов типа «но», «однако». Если во второй части предложения не обнаружено оценочных слов, но присутствуют союзы, то второй части предложения должна быть приписана оценка, противоположная оценке первой части предложения.

В результате должно быть проведено агрегирование оценок по каждой аспектной категории. В статье [30] предложена следующая процедура проставления оценок аспектов в отдельном предложении. Пусть в предложении  $s$  содержится набор аспектных терминов  $\{a_1, \dots, a_n\}$  и оценочных выражений  $\{sw_1, \dots, sw_n\}$ , для которых оценки из словаря уже модифицированы с учетом операторов и контекста. Тогда оценки тональности каждого аспектного термина вычисляются по следующей формуле:

$$score(a_i, s) = \sum_{sw_j} \frac{sw_j so}{dist(sw_j, a_i)},$$

где  $sw_j$  – оценочное слово или выражение,  $sw_{jso}$  – числовая оценка тональности  $sw_j$ ,  $dist(sw_j, a_i)$  – расстояние между оценочным словом и аспектом. Таким образом, к каждому аспектному термину в предложении приписываются все оценки, упомянутые в этом предложении, однако их вес падает в зависимости от расстояния между аспектом и оценкой. Если окончательный вес положительный, то и оценка аспекта положительная, отрицательный вес означает отрицательную оценку, вес 0 – нейтральную оценку.

Авторами в работе [30] используются шесть правил композиции оценок для определения тональности по отношению к объекту: конверсия тональности, агрегация, распространение, доминирование, нейтрализация и интенсификация.

Конверсия – это применение отрицаний и перевод в противоположную тональность. Агрегация применяется для синтаксических групп вида прилагательное-существительное, существительное-существительное, наречие-прилагательное, наречие-глагол, имеющих противоположную тональность, например, *«прекрасная битва»*. В таком случае этой фразе приписывается доминирующая тональность модификатора:  $POS(«прекрасная») \& NEG(«битва») \Rightarrow POSneg(«прекрасная битва»)$ .

Правило распространения применяется, когда в предложении употребляется глагол распространения или передачи:  $PROP-POS(«восхищаться») \& «его поведением» \Rightarrow POS(«его поведением»)$ .

Правило доминирования заключается в том, что если полярности глагола и его объекта различны, то полярность глагола преобладает  $NEG(«обмануть») \& POS(«надежды») \Rightarrow NEG(«обмануть надежды»)$ . Если в сложном предложении фразы соединены союзом «но», то тональность второй части предложения доминирует: например:  $NEG(«Было тяжело карабкаться на гору всю ночь»), but POS(«открывшийся волшебный вид стал наградой для путешественников») \Rightarrow POS(предложение)$ .

Правило нейтрализации применяется, когда предлог-модификатор или оператор условия относится к тональному выражению, например, «*несмотря на*» & NEG(«*переживания*») => NEUT(«*несмотря на переживания*»). Правило интенсификации усиливает или ослабляет вес тональности, например, POS(«*счастливый*») < POS(«*очень счастливый*»).

## **2.3.2. Машинное обучение**

В машинном обучении определение тональности почти всегда решается как задача классификации (обучение с учителем). Для этого требуется заранее размеченная выборка. Конкретно глубокое обучение при этом отличается лишь менее интуитивными моделями, которые могут выстраивать более сложные зависимости между словами и их признаками.

### **2.3.2.1. Стандартные методы машинного обучения**

Результаты применения методов машинного обучения в высокой степени зависят от обучающей выборки – модель, настроенная на одних размеченных данных часто показывает заметно худшее качество при работе в другой предметной области. Хотя эта проблема известна давно и существуют исследования в этом направлении, предложенные идеи пока еще плохо работают на практике. В этом заключается основной недостаток подхода, связанного с машинным обучением.

Glavaš и др. [7] на основе заранее собранных оценочных слов и аспектов рассматривают задачу классификации, т. е. для заданного предложения классификатор должен проставить, к какому именно аспектному термину относится данное оценочное слово, что может быть существенным для длинного предложения, в котором упомянуто несколько оценок и несколько аспектов.

Задача классификации в машинном обучении довольно хорошо изучена и предложен широкий выбор готовых алгоритмов для практического использования. Конкретно в контексте определения тональности суждений исследователями часто применяются методы SVM, KNN, логистическая регрессия, наивный байесовский классификатор.

В качестве признаков для классификации в рамках машинного обучения обычно рассматриваются:

- признаки расположения: расстояние между аспектным термином и оценочным словом, число аспектов и оценочных слов в предложении, длина предложения, пунктуация, наличие одних аспектов между другими аспектами и оценочными словами, порядок расположения аспекта и оценочного слова,
- лексические признаки: набор слов между аспектным термином и оценочным словом, наличие союзов, междометий, смайликов, хэштегов, восклицаний и др.,
- морфологические признаки: время, наклонение, вид, возвратность,
- части речи оценочного слова и аспектного термина, набор тегов частей речи между аспектом и оценочным словом, части речи соседних слов,
- признаки, основанные на синтаксической структуре: набор тегов по пути между аспектом и оценочным словом, близость по синтаксическому дереву.

Что касается преобразования слов в форму, пригодную для модели машинного обучения, то наиболее простой подход – использование «мешка слов» или «мешка N-граммов». В первом варианте, например, рассчитывается количество вхождений каждого слова в предложение. Такой подход в основном не учитывает контекст и семантику языка, но часто дает вполне неплохие результаты на практике. Признаками в данной ситуации выступают векторы из слов заранее заданной длины. Таким образом, происходит потеря части информации о близости слов в предложении, что может являться важным признаком. В последнее время, однако, почти всегда применяется подход по типу Word2Vec или Glove, так как закодированное в вектор слово уже учитывает контекст.

#### **2.3.2.2. Глубокое обучение**

В статье [40] авторы использовали те же архитектуры нейронных сетей, которые применили ранее для поиска аспектных терминов. Примечательно, что была обнаружена особенность рекуррентных моделей выучивать отрицания в тексте без введения специальных лингвистических правил и шаблонов (например, словосочетание *«не нравится»*). Авторы также заметили, что использование скрытого слоя RNNLM вместо векторного представления слов может повысить качество модели, так как лучше учитываются связи в предложении. Лучшей моделью оказалась LSTM, обученная на таких признаках.

Karпов I. и др. [15], оценивая тональность сообщений в Twitter, сравнивали подход, основанный на синтаксических правилах, с моделью сверточной нейронной сети. Последняя также использовала информацию о синтаксических связях наряду с линейным порядком слов в предложении. Авторы специально подобрали фильтры свертки для CNN таким образом, чтобы они были похожи на отдельные лингвистические правила из первого метода. Результаты показали, что CNN с использованием Word2Vec векторов, полученных на корпусе текстов Twitter, значительно превосходит



по точности традиционный подход. Гибридная модель не дает каких-либо заметных преимуществ.

Аналогичные данные в рамках конкурса SentiRuEval были использованы другой командой участников [1]. Сравнивая, между собой архитектуры CNN, LSTM и GRU, авторы пришли к выводу, что наиболее оптимальной моделью является GRU с 2 слоями. Выбранная нейросеть заметно превзошла по качеству CNN и показала немного лучшие результаты, чем LSTM. Интересно, что использование последовательности слов в обратном порядке позволило еще повысить качество классификации. Также была сделана попытка построить ансамблевый классификатор из LSTM, GRU и SVM с полиномиальным ядром. Однако, ансамблевый метод оказался несколько хуже выбранной нейросети.

В работе [13], упоминавшейся ранее, авторы также использовали GRU с обратными связями и полносвязным слоем для бинарной классификации на положительную или отрицательную тональность. Пространство признаков было получено с помощью конкатенации векторов слов, понятий, частей речи и расстояния от аспекта. Для определения принадлежности к конкретному понятию использовался ресурс SenticNet. Расстояние до аспекта выражалось также в форме вектора, пространственная размерность которого являлась гиперпараметром модели. Сама идея использования расстояния связана с общей интуицией о том, что ближайший контекст имеет большее влияние на целевое слово. Альтернативный метод учета расстояния – использование деревьев синтаксического разбора [43] с нормализацией на длину предложения. Другой подход – использование окна заданной ширины проще в реализации, но при этом менее информативен.

Irsoy O. и Cardie C. [10] проводят подобное сравнение архитектур RNN. Авторы зафиксировали статистически значимое улучшение качества с использованием моделей с обратными связями. Анализируя результаты нейронных сетей с различным количеством скрытых слоев, было обнаружено, что увеличение «глубины» модели до определенной степени

имеет позитивный эффект. Оптимальное число скрытых слоев RNN – 2 или 3, в зависимости от данных.

Исследователи в статье [24] подробно описывают процесс сравнения различных типов рекуррентных нейросетей между собой и CRF. Во всех случаях RNN – модели показали существенно лучший результат. Elman-type RNN оказалась точнее Jordan-type RNN, при этом уступив LSTM. В отличие от работы [10] авторы отказались от использования обратных связей, т.к. это всегда ухудшало результат. Одно из объяснений – отличие в разметке данных. Применение информации о принадлежности к частям речи позволило добиться весьма заметного улучшения качества.

В аспектном анализе часто возникает вопрос как связать между собой этапы извлечения аспектных терминов, их группировку и определение тональности полученной аспектной категории. Так, например, в статье [45] предлагается усреднить в предложении векторы аспектов, принадлежащих к одинаковой группе. Полученное представление аспектной группы совместно с остальными векторами слов авторы впоследствии применяют в модели CNN.

## **2.4. Извлечение прочих сущностей текста**

Помимо перечисленных выше сущностей, в реальных задачах иногда важно определить главный объект в тексте и получить информацию об авторе и времени. Однако, несмотря на практическую важность данного этапа, он останется за пределами данной работы. Во-первых, используемые для экспериментов данные не содержат информацию о времени и авторе. Во-вторых, такие сущности часто уже содержатся в заголовках текстов и их извлечение не представляет собой особого научного интереса.

Для извлечения объектов пользователь может заранее составить список интересующих объектов, в частности, названия фирм-конкурентов. Однако, такой список не будет исчерпывающим, так как невозможно предсказать появление новых объектов, а так же охватить все синонимы.

В случае, когда объект не задан явно, проблема становится стандартной задачей извлечения именованных сущностей или информации. Самые ранние идеи основываются на лингвистических правилах и шаблонах, и описаны, например, в работе [34]. Более современные статистические методы часто используют ранее упомянутые HMM и CRF.

В исследовании [22] задача поиска объектов сформулирована следующим образом: для набора  $Q$  базовых объектов класса  $C$  и множества объектов-кандидатов  $D$ , необходимо определить, какие из элементов  $D$  принадлежат  $C$ . В результате класс  $C$  «наращивается» из базового множества  $Q$ . Формально, задача сводится к стандартной классификации или ранжированию. Классический подход – сравнение схожести ближайшего контекста каждого объекта-кандидата с явно заданными первоначальными объектами.

Для извлечения авторов (субъектов) в текстах Choi и др. [4] использовали CRF с контекстными признаками – смежные слова и их части речи, грамматические роли, слова тональности и пр. В работе [16] исследователи сначала генерируют список всех возможных субъектов для предложения, например, используя только именные группы, местоимения и т.п. После этого строится модель максимизации энтропии с признаками, полученными из дерева разбора предложения. Модель производит ранжирование кандидатов с последующим выбором слова с наивысшей оценкой. Johansson и др. [14] использовали стандартный метод SVM.

В исследовании [35] авторы предлагают использовать автоматическую разметку семантических ролей (ASRL) для извлечения субъектов в тексте. Однако, этот подход обозначен как несамодостаточный и следует дополнительно учитывать структуры дискурса.

## 3. Проведение и анализ экспериментов

### 3.1. Описание данных

Проведение подобных исследований неизбежно связано с проблемой ограниченности данных для обучения модели. Разметка коллекции текстов для аспектного анализа – долгий и трудоемкий процесс, в отличие, например, от разметки изображений. Дополнительная сложность заключается в том, что, например, в некоторых ситуациях не существует окончательно верного варианта определения тональности по отношению к объекту. Человек может воспринимать одну и ту же информацию по-разному. Правильное с точки зрения качества разметки решение – учитывать несколько экспертных мнений в ситуации неопределенности, что, однако, увеличивает время и стоимость получения данных.

Проблема аспектного анализа приобрела популярность лишь в последние несколько лет. Основные работы по этой тематике написаны по результатам экспериментов, проводимых в рамках ежегодных конкурсов SemEval. Российский аналог – конкурс SentiRuEval, в котором проводится анализ текстов также и на русском языке.

#### 3.1.1. Данные SentiRuEval

Одна из секций конференции SentiRuEval-2015 была посвящена задаче аспектного анализа на основе отзывов интернет-пользователей. Организаторы предложили участникам 2 коллекции текстов – о ресторанах и автомобилях [39]. Для каждой группы отзывов заранее был определен список аспектов. Так, например, для ресторанной тематики:

- еда;
- сервис;
- интерьер;
- цена;
- общий аспект (объект в целом).

Для текстов про автомобили:

- управляемость;
- надежность;
- безопасность;
- внешний вид;
- комфорт;
- стоимость;
- общий аспект.

Каждый целевой аспект может выражаться в тексте набором аспектных терминов. Как уже было сказано выше, каждый термин может быть одним из нескольких типов:

- явный («*Меню достаточно разнообразное и весьма вкусное*»);
- неявный («*Этот автомобиль очень дорогой*»);
- тональный факт («*В салате было много майонеза*»).

Аспектный термин может обладать одним из 4 видов тональности:

- позитивная;
- негативная;
- противоречивая;
- нейтральная (отсутствие выраженной тональности).

Общая информация о коллекциях размеченных текстов указана в Таблице 1.

*Таблица 1. Статистика размеченных данных SentiRuEval.*

Количество в выборке	Рестораны		Автомобили	
	Обучение	Тест	Обучение	Тест
отзывов	201	203	217	201
явных аспектов	2,822	3,506	3,152	3,109
неявных аспектов	636	657	638	576
тональных фактов	523	656	668	685

аспектов с позитивной тональностью	2,530	3,424	2,330	2,499
аспектов с негативной тональностью	684	865	1,337	1,300
аспектов с нейтральной тональностью	714	445	691	456
аспектов с противоречивой тональностью	53	85	100	115

Из данных табл. 1 видно, что представленная выборка текстов сравнительно небольшая по размеру и разделена в примерной пропорции 1:1 на тестовые и обучающие отзывы. Позитивный факт заключается в том, что большинство аспектных терминов относится к типу явных, которые алгоритм проще идентифицирует.

Важно отметить, что корпус текстов подбирался не случайным образом [26], следовательно нельзя напрямую примерять полученные результаты экспериментов к реальным жизненным задачам. Организаторы SentiRuEval использовали отзывы примерно одинаковой длины с похожим содержанием. Для этого применялся простой подход «мешка слов»: на большой коллекции интернет-отзывов был рассчитан усредненный вектор вхождений слов, отражающий черты некоторого типичного отзыва. Для финальной выборки отбирались те тексты, вектор «мешка слов» которых находился ближе всего к усредненному вектору с точки зрения косинусного расстояния.

Одно из главных очевидных последствий применимого метода – все выбранные тексты должны обладать похожим словарем. Это, с одной стороны, позволяет уменьшить размер достаточной выборки для обучения моделей, и следовательно, сократить организационные издержки. С другой стороны, это увеличивает необъективность оценки качества результатов. Вполне вероятно, что лучшая выбранная в рамках конкурса модель окажется неоптимальной на тестовой выборке из случайно отобранных отзывов.

### 3.1.2. Дополнительные данные

Кроме официальных данных SentiRuEval, было использовано 3 больших неразмеченных русскоязычных корпуса текстов, статистика по которым представлена в Таблице 2:

Таблица 2. Дополнительные данные для экспериментов.

№ п/п	Русскоязычный корпус	Приблизительное кол-во слов
1.	Интернет-новости (09.2013 – 11.2016)	5 млрд
2.	Отзывы об автомобилях, в т.ч.	99 млн
	<i>drom.ru</i>	48 млн
	<i>cars.mail.ru</i>	35 млн
	<i>my.auto.ru</i>	16 млн
3.	Отзывы о ресторанах, в т.ч.	254 млн
	<i>tripadvisor.ru</i>	122 млн
	<i>afisha.ru</i>	94 млн
	<i>allcafe.ru</i>	38 млн

Дубликаты отзывов из различных интернет-ресурсов были удалены. Из представленной статистики ясно, что новостная коллекция текстов значительно превосходит каждый из корпусов по заданной тематике.

Потребность в использовании дополнительных корпусов текстов связана с получением векторов слов в модели Word2Vec. Отзывы рассматриваемой предметной области часто содержат специфические термины или слова, которые могут встречаться довольно редко в обычных текстах.

## 3.2. Постановка задачи и метрики качества

В рамках проведения конкурса SentiRuEval–2015 участникам были предложены следующие задания:

- Задание А: извлечение только явных аспектных терминов.
- Задание Б: извлечение всех аспектных терминов.
- Задание В: определение тональности аспектных терминов.
- Задание Г: определение аспектной группы для явных терминов.
- Задание Д: определение интегральных оценок.

Задание Б подразумевает извлечение неявных аспектов и тональных фактов в дополнение к явным аспектам. Задания В и Д используют одинаковую шкалу из 4 значений, которая упоминалась ранее. В случае отсутствия в тексте аспектных терминов определенной категории, оценка категории не проставляется. При этом оценка по объекту в целом должна быть выставлена в любом случае. Если объект обладает положительной тональностью хотя бы для одной категории людей, то отзыв считается положительным. Например, ресторан может рекомендоваться к посещению только для молодежи.

Для оценки результатов первых двух заданий организаторы использовали F1–меру, рассчитанную макро-способом, т.е. усреднением индивидуальных F1–мер по каждому отзыву.

Аспектный термин иногда состоит из нескольких слов, а модель может найти только его часть. Поэтому F1–мера рассчитывалась в строгом варианте полного соответствия аспекта предсказанию и частичном варианте. Для последнего применялись следующие формулы расчета точности (Precision) и полноты (Recall):

$$Precision_t = \frac{|t \cap t_s|}{|t_s|},$$

$$Recall_t = \frac{|t \cap t_s|}{|t|},$$



где  $t_s$  – предсказанный моделью аспектный термин. Адаптированные для данной задачи формулы соответствуют исходной логике расчета точности и полноты. Так, чтобы иметь высокую точность по вышеуказанной формуле, модели можно быть консервативной – лучше находить лишь часть аспектных терминов и не ошибаться, чем извлекать длинные составные термины и ошибаться. Для полноты справедлива обратная логика.

В задании классификации В использовалась F1–мера, рассчитанная макро– и микро–способами. В первом случае метрика вычислялась для каждого класса тональности в отдельности и затем усреднялась. В такой ситуации классы имеют равный вес. Микро–способ рассчитывается на матрице ошибок для всех классов, что приводит к увеличению влияния наиболее представленного класса тональности (положительная тональность).

Для последних двух заданий применялась только макро F1–мера.

В экспериментальной части данной работы использовались аналогичные метрики с целью сравнения полученных результатов с итогами конкурса SentiRuEval.

### 3.3. Базовое решение

Для каждого задания SentiRuEval, организаторами были предложены простые *baseline* решения с оценкой соответствующих метрик качества.

В заданиях извлечения аспектных терминов (А и Б) применялся следующий подход. На основе обучающей выборки создавался список всех размеченных аспектных терминов, затем проводилась их лемматизация. В тестовой коллекции извлекались все термины, которые совпали с леммами из списка. При своей простоте метод имеет ряд очевидных недостатков. Во-первых, необходима действительно большая обучающая выборка, чтобы учесть при обучении все возможные аспектные термины и их синонимы, что представляется невозможным. Во-вторых, неявные термины и тональные факты с большой вероятностью не будут найдены в тестовых текстах, т.к. они часто представляют собой уникальные сочетания слов, для которых никак нельзя заранее создать список. Поэтому, такой подход малоприменим на практике в решении бизнес-задач.

Для задания классификации (В, Г, Е) baseline-решение было еще более простым, которое, однако, показало себя лучше многих сложных систем. Например для задания В, тональность аспектного термина определялась как наиболее частотная эмоциональная окраска данного термина в обучающей коллекции. В случае если в обучающей выборке такого термина не было, то предсказывался самый распространенный класс тональности. Аналогичная процедура применялась для отнесения аспектного термина к определенной аспектной группе (Г) и определения интегральных оценок (Е). В случае последнего, алгоритм выглядит даже несколько абсурдным, т.к. общая тональность текста будет назначена как самая частотная общая тональность отзывов обучающей выборки.

Высокие результаты такого простого решения, как в заданиях В-Г можно объяснить тем фактом, что классы тональностей несбалансированы, т.е. положительная эмоциональная окраска выраженно преобладает над остальными. Поэтому простая стратегия – во всех неизвестных ситуациях предсказывать позитивную тональность – в данном конкретном случае может быть выигрышным решением. Эта стратегия, однако, не сработает в общем случае, поэтому на практике метод также не применим. Например, в ресторане сменился шеф-повар и отзывы посетителей о кухне резко сменились с положительных на отрицательные. Алгоритм, однако, будет продолжать давать позитивные предсказания тональности, т.к. они были преобладающими в обучающей выборке.

### 3.4. Модели для экспериментов

Экспериментальная часть данной работы следует общей схеме системы аспектного анализа, которая представлена на рисунке 2:

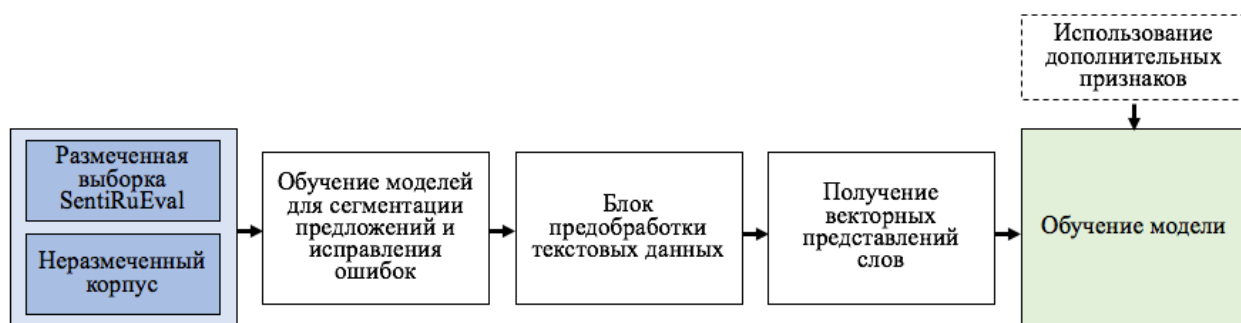


Рисунок 2. Схема проведения экспериментов.

Коллекция неразмеченных текстов была представлена одним из 2 корпусов:

1. Интернет-новости.
2. Отзывы заданной тематики (автомобильные или ресторанные).

Далее вся коллекция текстов использовалась для обучения алгоритма сегментации предложений с помощью модуля Punkt библиотеки NLTK, который представляет собой имплементацию алгоритма обучения без учителя [17]. Чтобы правильно найти границы предложений в тексте, алгоритм учитывает сокращения, устойчивые словосочетания, слова с которых часто начинаются предложения. Сегментированные предложения используются Word2Vec для получения векторов слов.

Выбранный корпус также применялся для обучения модели исправления ошибок/опечаток в текстах отзывов. Основная идея базируется на использовании небольшого набора правил, отражающих наиболее частые неточности написания – пропуски, удвоения, смена мест букв и т.п. Используя эти правила, алгоритм создает список возможных исправлений слова и выбирает среди них вариант, который наиболее часто встречается в исходном корпусе.

Блок предобработки текстов – важная часть анализа. Он состоит из следующих этапов:

- сегментация предложений (на основе обученной шагом ранее модели);
- исправление ошибок (на основе обученной шагом ранее модели);
- приведение слов к нижнему регистру;
- лемматизация слов;
- замена чисел на специальный символ NUM;
- удаление пунктуации.

Опциональные этапы предобработки текстов, влияние которых на качество модели будет показано отдельно:

- учет символов эмоций (EMOJ\_POS, EMOJ\_NEG);
- учет пунктуации (EXCL, QUEST, COMM, COLON).

Получение векторных представлений слов осуществлялось с помощью модели Word2Vec, о которой будет рассказано подробнее в следующем пункте. В данной работе применялась имплементация алгоритма библиотеки Gensim.

В качестве дополнительных признаков для обучения использовалась информация о принадлежности к частям речи и морфологический признак времени для глаголов. В обоих случаях применялась библиотека `rumorphy2`.

### 3.4.1. Word2Vec

Алгоритм Word2Vec представляет собой популярный метод получения векторных представлений слов, описанный в работе Mikolov T. [29]. Суть метода заключается в попытке смоделировать контекст выбранного слова. Для этого используется один из двух методов:

1. Continuous bag-of-words (CBOW).
2. Skip-gram.

В первом случае модель обучается предсказывать слово по его ближайшему контексту, во втором – наоборот. Для дальнейших экспериментов будет использован второй метод, т.к. по утверждениям самих авторов, он больше подходит для небольших коллекций текстов, каковыми являются корпуса отзывов.

Архитектура Word2Vec представляет собой нейронную сеть с одним скрытым слоем, изображенную на рисунке 3. Входящий вектор и слой с предсказанными вероятностями имеют одинаковую размерность, равную длине определенного словаря наиболее частотных слов. Для получения векторного представления слов используются нейроны скрытого слоя.

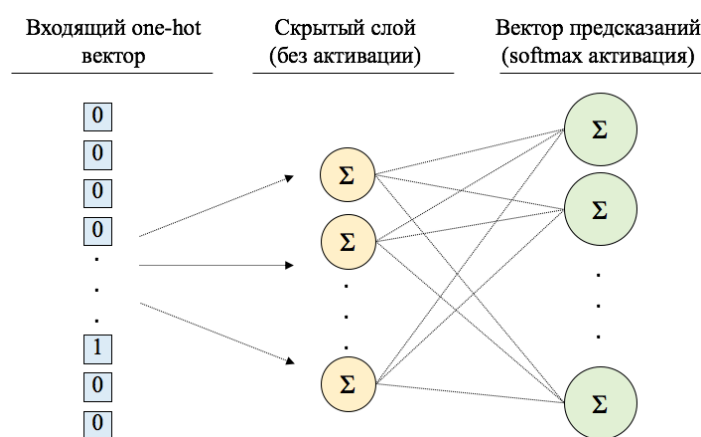


Рисунок 3. Skip-gram реализация Word2Vec.

Модель Word2Vec обладает рядом гиперпараметров, которые могут быть подобраны экспериментальным методом. Для оценки качества получившихся векторов существует ряд специфических задач (поиск лишнего слова, заполнение пропусков и т.п.), доступных, однако, только для модели, обученной на корпусе текстов английского языка. Поэтому для подбора оптимальных параметров был использован простой перебор всех основных комбинаций, использованных ранее авторами в похожих работах. В результате этого, финальная модель обладает следующим набором параметров:

- размерность векторов – 300;
- ширина окна – 5;
- минимальный порог вхождения слова – 3,
- коэффициент `downsampling` – 0.001 (снижение частоты частотных слов).

### **3.4.2. Сверточные нейронные сети**

Архитектура сверточной нейросети была предложена изначально для решения задач распознавания образов. Однако, позже выяснилось, что данная модель также хорошо справляется с некоторыми задачами автоматической обработки текстов. Например, для решения проблемы определения тональности, CNN используется практически в половине работ, построенных на применении глубокого обучения.

Что касается задач разметки последовательностей (в данном случае, поиск аспектных терминов), то лишь немногие исследователи используют сверточные нейронные сети. Во-первых, это связано со сложностями в обучении модели, так как архитектура CNN по умолчанию не устроена для решения подобных задач. Во-вторых, качество такой нейросети часто проигрывает главным конкурентам – CRF и RNN. Поэтому, в данной работе CNN будет использоваться для целей классификации (задания В–Д SentiRuEval).

Общий вид архитектуры CNN, применимой для NLP-задач, показан на рисунке 4:

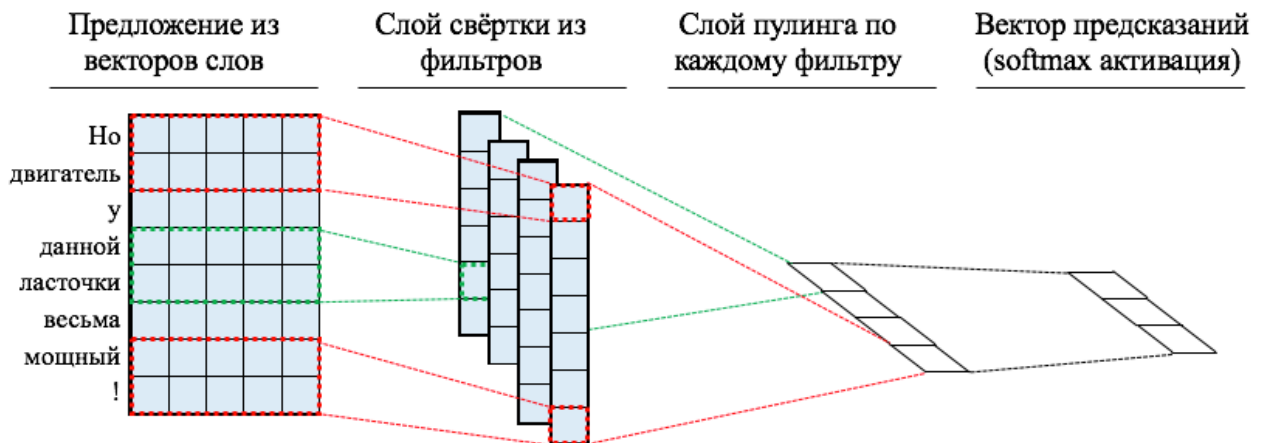


Рисунок 4. CNN с одним слоем свёртки и пулинга.

Сверточная нейронная сеть принимает на вход матрицу предложения, состоящую из конкатенированных векторов слов, полученных с помощью Word2Vec. Сверточный слой состоит из набора фильтров заданной ширины, другая размерность которых совпадает с размерностью векторов слов. Результат свертки получается путем простого поэлементного умножения части входной матрицы. Пулинг «собирает» результат с каждого фильтра, находя максимум значений из соответствующих значений. Слой пулинга трансформируется в вероятности предсказанных классов с помощью softmax активации:

$$\varphi(x)_j = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}},$$

где  $K$  — количество классов.

Для определения тональности аспектного термина, на вход CNN подается предложение, содержащее данный термин. На выходе рассчитываются вероятности каждого из 4 классов тональностей. Дополнительно был учтен позиционный признак расстояния каждого слова от текущего аспектного термина в соответствии с методикой, предложенной в [38]. Данный признак позволяет алгоритму понять, какой аспект рассматривается в настоящий момент в предложении с несколькими аспектными терминами.

Аналогичный метод использован для определения принадлежности термина к аспектной категории и построения интегральных оценок категорий. Тональность объекта в целом, при отсутствии упоминания в тексте, предсказывалась по всему отзыву.

Результаты экспериментов показали, что сверточные нейросети существенно хуже справляются с задачами классификации. Поэтому в таблице с результатами отражено качество только лучшей выбранной архитектуры CNN:

- слой свертки 10 фильтров шириной 2 и шагом 1 (ReLU активация);
- слой макс-пулинга с размером окна 2;
- слой свертки с 10 фильтрами шириной 3 и шагом 1 (ReLU активация);
- слой макс-пулинга по оси времени;
- слой *dropout* с вероятностью 0.2;
- выходной слой (softmax активация).

Поиск оптимальных значений параметров нейронной сети осуществлялся с помощью оптимизационного метода *Adam*. Для контроля переобучения использовалась кросс-валидация с разделением выборки на 5 частей. Целевая функция оптимизации для CNN и RNN – категориальная кросс-энтропия:

$$L_i = - \sum_j t_{i,j} \log(p_{i,j}),$$

где  $t_{i,j}$  – метка  $i$ -го наблюдения для класса  $j$ ,  $p_{i,j}$  – предсказанная вероятность класса  $j$  для  $i$ -го наблюдения.



### 3.4.3. Рекуррентные нейронные сети

Под общим названием рекуррентных нейронных сетей обычно понимают семейство похожих архитектур, главные из которых – RNN, GRU, LSTM. Основная общая черта перечисленных методов – работа с последовательностями и решение таких типовых задач как генерация описания изображений, последовательностей ответов на вопросы (чат-боты), разметка частей речи в предложении и т.д. Важной встроенной особенностью является способность создавать предсказания как на каждом шаге последовательности (принцип «many to many»), так и после прочтения всей последовательности («many-to-one»):

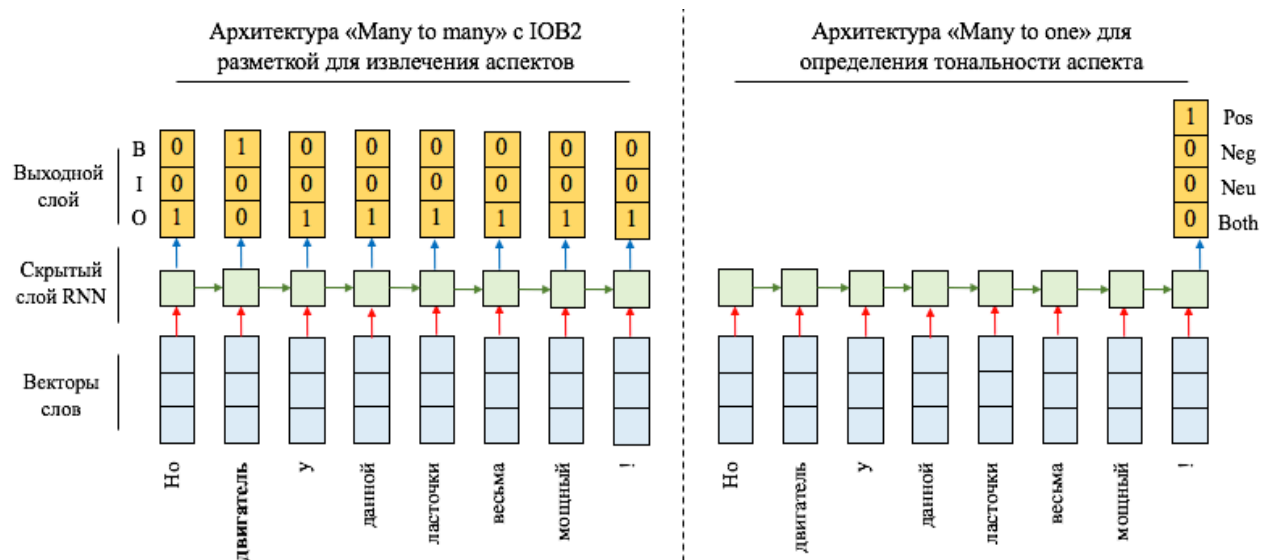


Рисунок 5. Архитектура RNN в зависимости от задачи.

На вышеприведенном рисунке показана архитектура рекуррентной нейронной сети с одним скрытым слоем для задачи извлечения аспектных терминов (слева) и задачи классификации (задания В-Г) справа. Для задачи поиска аспектов применяется распространенная кодировка IOB2, представляющая собой one-hot вектор из 3 элементов. Начало термина обозначается буквой *B*, продолжение – *I*, во всех остальных случаях – *O*. На каждом шаге модель получает новое слово-вектор, комбинирует его с прошлым скрытым состоянием и делает прогноз для текущего слово, является ли оно аспектом.

Для задачи классификации нейронная сеть предсказывает значение класса только на последнем шаге после «прочтения» всей последовательности.

Существует несколько известных вариаций RNN, с которыми далее будут проведены эксперименты. Самая первая и популярная конфигурация – *Elman-type RNN*, в которой текущее скрытое состояние  $h_t$  является нелинейной трансформацией прошлого состояния и текущего наблюдаемого значения  $x_t$ :

$$h_t = f(Uh_{t-1} + Vx_t + b),$$

где  $f$  – функция активации,  $b$  – вектор сдвига,  $U$  и  $V$  – матрицы весов.

Скрытое состояние  $h_t$  можно интерпретировать как накопленные знания о прошлом, которые могут накапливаться и меняться во времени.

Немного другая архитектура – *Jordan-type RNN*. Отличие состоит в том, что скрытое текущее состояние зависит от предсказанного значения на прошлом шаге  $y_{t-1}$ :

$$h_t = f(Uy_{t-1} + Vx_t + b).$$

В работе с текстовыми последовательностями иногда важно учитывать не только прямой, но и обратный порядок слов, т.к. знание о будущем может содержать ценную информацию. Для этого на практике применяется архитектура RNN с обратными связями (*bidirectional RNN*). Матрицы весов  $U$  и  $V$ , а также сдвиг  $b$  и скрытые состояния  $h$  независимы для слоев с прямыми и обратными связями. Предсказания создаются путем комбинации 2 скрытых состояний:

$$y_t = g(\vec{W}\vec{h}_t + \overleftarrow{W}\overleftarrow{h}_t + c).$$

Типы рекуррентных нейронных сетей, названные выше, изображены на рисунке 6:

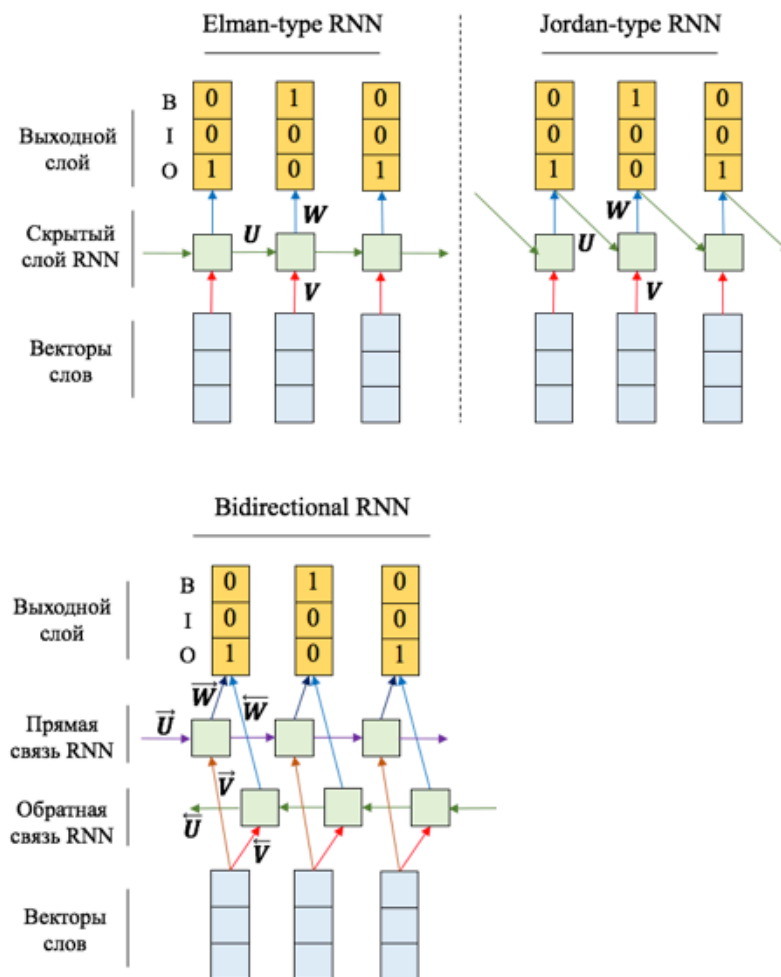


Рисунок 6. Основные типы RNN.

Использование RNN на практике связано с известной проблемой затухания градиента (*vanishing gradient problem*) при обучении нейронной сети методом обратного распространения ошибки. Проблема возникает из-за многочисленных звеньев в поиске градиента сложной функции. Одно из неприятных последствий – неспособность RNN хранить информацию о длинных последовательностях. Для разрешения ситуации исследователями была предложена архитектура LSTM (*Long short-term memory*) и более современная упрощенная альтернатива – GRU (*Gated recurrent unit*).

Основное отличие LSTM и GRU от RNN заключается в измененном методе расчета активаций скрытого слоя. Скрытое состояние LSTM на текущем шаге  $h_t$  обычно задается набором из следующих уравнений [8]:

$$\begin{aligned} i_t &= \sigma_i(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \\ f_t &= \sigma_f(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \\ c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \\ o_t &= \sigma_o(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \\ h_t &= o_t \tanh(c_t), \end{aligned}$$

где  $\sigma$  – сигмоидная активация,  $\tanh$  – гиперболический тангенс,  $W$  – матрица весов,  $b$  – вектор сдвига,  $i$  – вектор *input gate*,  $f$  – вектор *forget gate*,  $c$  – вектор *cell*,  $o$  – вектор *output gate*.

Модель GRU является упрощением LSTM. GRU обладает меньшим числом параметром, что положительно сказывается на скорости обучения, при сопоставимом качестве результатов. Активация скрытого слоя рассчитывается по формулам [3]:

$$\begin{aligned} r_t &= \sigma_r(W_{xr}x_t + W_{hr}h_{t-1} + b_r), \\ u_t &= \sigma_u(W_{xu}x_t + W_{hu}h_{t-1} + b_u), \\ c_t &= \sigma_c(W_{xc}x_t + r_t \odot W_{hc}h_{t-1} + b_c), \\ h_t &= (1 - u_t) \odot h_{t-1} + u_t \odot c_t, \end{aligned}$$

где  $\odot$  означает поэлементное умножение,  $r$ ,  $u$ ,  $c$  – векторы *reset gate*, *update gate*, *cell memory* соответственно.

Для обучения рекуррентных нейросетей использовалась аналогичная процедура, описанная в предыдущем разделе. Регуляризация *drop-out* применялась в соответствии с методикой, изложенной в работе [6]. Были проведены эксперименты с размерностью скрытого слоя (от 50 до 300 нейронов с шагом 50) и количеством скрытых слоев. Наиболее оптимальные комбинации гиперпараметров представлены в таблицах с результатами ниже.

### 3.5. Оценка результатов

Ниже рассмотрены результаты полученных экспериментов в соответствии с заданиями и метриками SentiRuEval. Эксперименты проводились в основном для рекуррентных нейросетей, так как CNN ограничена в решении задач извлечения аспектов, а для определения тональности показывает заметно худшее качество. Поэтому для сравнения была выбрана только лучшая архитектура.

Кроме базового решения, предложенного организаторами SentiRuEval, для задачи извлечения аспектов приведены результаты использования CRF одним из участников. Аналогично, для заданий В-Д показаны результаты применения других стандартных алгоритмов машинного обучения – Gradient Boosting Classifier и SVM.

Для моделей с высокими результатами во втором пункте данного раздела применялся расширенный этап предобработки и дополнительный набор признаков.

#### 3.5.1. Выбор оптимальной модели

В таблице 3 представлены результаты выполнения задания по извлечению явных аспектных терминов из отзывов.

Можно с уверенностью сказать, что векторные представления слов, полученные на корпусе текстов исследуемой тематики, ощутимо улучшают качество работы алгоритма. Особенно этот факт проявляется для автомобильных отзывов, вероятно, из-за более выраженной специфики словаря. В целом, при оптимально выбранной архитектуре, нейронные сети показывают весомо лучший результат, чем альтернативный метод CRF. Наличие у рекуррентной нейросети слоя с обратными связями скорее снижает качество модели, возможно, из-за переобучения ввиду большого количества параметров.

LSTM и GRU конфигурации практически во всех случаях превосходят более простые альтернативы рекуррентных нейросетей. Оптимальное количество нейронов в скрытом слое обратно пропорционально сложности модели. Elman-RNN обладает более высокими значениями метрик качества по сравнению с Jordan-RNN.

Таблица 3. Результаты извлечения явных аспектных терминов.

Корпус Word2Vec	Кол-во нейронов	Автомобили		Рестораны	
		F1 макро (полное совпад.)	F1 макро (частичное совпад.)	F1 макро (полное совпад.)	F1 макро (частичное совпад.)
Базовое решение					
-	-	59.41	69.66	60.84	66.51
CRF участников SentiRuEval					
-	-	62.56	69.91	58.99	65.20
Elman-type RNN					
Новости	200	58.34	67.89	57.13	66.70
Отзывы	250	60.08	69.51	59.15	67.02
Bidirectional Elman-type RNN					
Новости	100	57.93	67.82	56.72	66.34
Отзывы	150	60.29	69.47	58.95	66.40
Jordan-type RNN					
Новости	200	58.01	67.12	56.83	66.14
Отзывы	250	60.02	69.58	57.88	66.18
GRU					
Новости	150	62.87	69.34	60.36	67.18
Отзывы	150	64.71	72.24	61.14	67.44
Bidirectional GRU					
Новости	100	62.43	69.32	59.41	66.94
Отзывы	150	64.52	71.91	60.48	67.02
LSTM					
Новости	100	64.51	70.47	61.18	67.44
Отзывы	150	66.94	72.18	62.19	68.49
Bidirectional LSTM					
Новости	50	64.55	70.19	61.01	67.13
Отзывы	100	66.50	71.84	61.99	68.36

Для задания Б (извлечение *всех* аспектных терминов, в т.ч. неявных и тональных фактов) общие результаты выглядят очень похоже. Соответствующая таблица представлена ниже:

*Таблица 4. Результаты извлечения всех аспектных терминов.*

Корпус Word2Vec	Кол-во нейронов	Автомобили		Рестораны	
		F1 макро (полное совпад.)	F1 макро (частичное совпад.)	F1 макро (полное совпад.)	F1 макро (частичное совпад.)
Базовое решение					
-	-	58.86	67.43	58.72	61.93
CRF участников SentiRuEval					
-	-	57.81	65.61	54.69	59.53
Elman-type RNN					
Новости	250	57.80	67.26	56.60	66.08
Отзывы	250	59.52	68.87	58.60	66.40
Bidirectional Elman-type RNN					
Новости	150	57.39	67.19	56.19	65.73
Отзывы	200	59.73	68.83	58.40	65.79
Jordan-type RNN					
Новости	250	57.47	66.50	56.30	65.53
Отзывы	300	59.46	68.94	57.34	65.57
GRU					
Новости	200	60.29	68.70	59.80	66.56
Отзывы	250	62.11	70.57	60.57	<b>66.82</b>
Bidirectional GRU					
Новости	100	60.35	68.51	59.14	65.92
Отзывы	150	61.87	69.98	60.08	66.22
LSTM					
Новости	100	61.91	69.82	60.28	66.01
Отзывы	150	<b>63.32</b>	<b>71.13</b>	<b>61.64</b>	66.55
Bidirectional LSTM					
Новости	100	61.09	69.73	60.24	65.52
Отзывы	100	62.54	70.56	60.97	66.01

В таблице 5 показаны результаты экспериментов для задания по определению тональности аспектных терминов. В дополнение к предыдущим моделям глубокого обучения также была обучена сверточная нейронная сеть, которая, однако, оказалась менее эффективной.



Значения микро F1–меры для автомобильной тематики существенно ниже. Это объясняется еще большим дисбалансом классов тональности для данных отзывов. В противоположность предыдущим выводам, использование корпуса отзывов для обучения Word2Vec не оказывает однозначного влияния на значения метрик качества. Методы глубокого обучения, особенно CNN, в данной задаче заметно менее эффективны, чем Gradient Boosting Classifier.

Таблица 5. Результаты определения тональности аспектных терминов.

Корпус Word2Vec	Кол-во нейронов	Автомобили		Рестораны	
		F1 micro	F1 macro	F1 micro	F1 macro
Базовое решение					
-	-	61.93	26.48	71.04	26.71
Gradient Boosting Classifier участников SentiRuEval					
-	-	74.28	56.84	82.49	55.46
CNN					
Новости	-	58.21	25.11	67.96	27.71
Отзывы	-	57.62	24.63	68.15	26.98
Elman-type RNN					
Новости	150	63.18	32.71	70.51	29.08
Отзывы	250	63.22	32.01	71.12	29.39
Bidirectional Elman-type RNN					
Новости	150	64.21	33.14	72.03	30.19
Отзывы	200	63.99	33.93	71.55	30.31
Jordan-type RNN					
Новости	200	62.16	32.78	70.44	28.23
Отзывы	300	62.01	32.54	71.13	28.81
GRU					
Новости	150	64.73	33.28	72.10	30.34
Отзывы	200	64.55	34.01	72.69	30.79
Bidirectional GRU					
Новости	100	65.84	33.68	73.52	32.12
Отзывы	100	65.91	34.04	72.90	31.74
LSTM					
Новости	100	64.98	33.09	71.98	32.03
Отзывы	100	65.14	33.58	72.15	31.67
Bidirectional LSTM					
Новости	50	66.17	34.78	73.66	32.14
Отзывы	100	65.43	35.18	73.01	31.96

Важно отметить, что архитектуры нейронных сетей с наличием обратных связей существенно эффективнее. Вероятно, для определения тональности аспектов важно учитывать последовательность после рассматриваемого слова. Полученные выводы применимы к заданию определения интегральных оценок (Таблица 7), где методы глубокого обучения также проигрывают по качеству SVM. В обоих заданиях лучшим вариантом нейронной сети является LSTM с обратными связями.

В задании Г может показаться, что алгоритмы лучше справляются с отзывами ресторанной тематики. Однако, этот факт может ввести в заблуждение, т.к. причина заключается лишь в большем количестве классов для автомобильных отзывов:

*Таблица 6. Результаты определения аспектной группы.*

Корпус Word2Vec	Количество нейронов	Автомобили	Рестораны
		F1 макро	F1 макро
Базовое решение			
-	-	56.36	79.96
SVM участников SentiRuEval			
-	-	65.21	86.53
CNN			
Новости	-	53.09	75.41
Отзывы	-	54.22	76.13
Elman-type RNN			
Новости	200	55.19	79.44
Отзывы	250	56.38	80.01
Bidirectional Elman-type RNN			
Новости	150	55.43	79.15
Отзывы	150	56.17	80.28
Jordan-type RNN			
Новости	200	55.09	78.73
Отзывы	250	55.88	79.25
GRU			
Новости	100	59.73	82.39
Отзывы	150	62.15	83.06
Bidirectional GRU			
Новости	100	58.47	81.96
Отзывы	100	60.81	82.14
LSTM			
Новости	100	60.13	82.90
Отзывы	150	61.69	83.72
Bidirectional LSTM			
Новости	50	59.13	81.12
Отзывы	100	61.23	82.35

Использование специализированного корпуса вносит положительный вклад в качество решения задачи. Так же, как и в прошлый раз, SVM выигрывает у методов глубокого обучения, наилучшие среди которых – GRU и LSTM.

*Таблица 7. Результаты определения интегральных оценок.*

Корпус Word2Vec	Кол-во нейронов	Автомобили F1 macro	Рестораны F1 macro
<b>Базовое решение</b>			
-	-	23.68	27.20
<b>SVM участников SentiRuEval</b>			
-	-	<b>43.90</b>	<b>45.81</b>
<b>CNN</b>			
Новости	-	25.01	26.99
Отзывы	-	26.32	27.12
<b>Elman-type RNN</b>			
Новости	200	26.91	28.74
Отзывы	200	26.12	27.39
<b>Bidirectional Elman-type RNN</b>			
Новости	150	28.43	31.63
Отзывы	200	28.70	30.84
<b>Jordan-type RNN</b>			
Новости	200	26.23	27.56
Отзывы	250	26.75	27.90
<b>GRU</b>			
Новости	150	29.14	32.62
Отзывы	150	29.45	32.01
<b>Bidirectional GRU</b>			
Новости	50	32.51	35.31
Отзывы	100	32.87	34.93
<b>LSTM</b>			
Новости	100	32.12	33.65
Отзывы	150	31.93	32.73
<b>Bidirectional LSTM</b>			
Новости	50	34.25	<b>36.20</b>
Отзывы	100	<b>34.92</b>	35.84

### 3.5.2. Использование дополнительных признаков

По итогам выполнения заданий SentiRuEval практически во всех случаях модель LSTM показала наилучший результат. В задачах определения тональности аспектных терминов в отдельности и аспектных групп в целом, было предпочтительней использовать LSTM с обратными связями. В остальных заданиях – LSTM с одним слоем последовательных связей. Для упрощения терминологии, под LSTM будет пониматься именно оптимальная, применительно к каждой конкретной задаче, архитектура модели.

В качестве расширенного этапа предобработки были учтены:

- эмоции (EMOJ\_POS, EMOJ\_NEG);
- пунктуация (EXCL, QUEST, COMM, COLON).

Сложность заключается в необходимости каждый раз заново обучать Word2Vec с каждой новой разметкой. Этот процесс занимает значительное время для новостного корпуса текстов. Поэтому для дальнейших экспериментов были использованы только отзывы рассматриваемой тематики.

Дополнительно были использованы признаки принадлежности к частям речи и время для глаголов:

- части речи (NOUN, ADJ, VERB, ADVB, PRT, GRND, NPRO, CONJ, PREP, PRCL, INTJ, OTHER);
- время (PAST, PRES, FUTR).

Идея использования времен глаголов может быть объяснена на примерах: «я не купил автомобиль» и «я не куплю автомобиль». В последнем случае прослеживается негативное высказывание автора по отношению к объекту, тогда как первая фраза, скорее всего, является лишь фактическим высказыванием.

Как было отмечено, учет дополнительных символов влияет на векторные представления слов и появление дополнительных токенов в последовательности. Добавление новых признаков было осуществлено с помощью обычной конкатенации векторов слов и *one-hot* векторов признаков.

Результаты экспериментов показаны в таблице ниже. В качестве упрощения анализа оценивались результаты по макро F1–мере, так как ее динамика практически всегда совпадает с микро F1–мерой. Для заданий извлечения аспектных терминов рассчитаны метрики полного соответствия.

Таблица 8. Результаты применения дополнительных признаков.

Модель	F1–macro				
	Задание А	Задание Б	Задание В	Задание Г	Задание Д
<b>Автомобили</b>					
LSTM	66.94	63.32	35.18	61.69	34.92
LSTM+Эм	66.17	62.98	36.02	61.84	36.11
LSTM+Пн	66.80	63.14	35.63	61.43	35.57
LSTM+Чр	<b>68.01</b>	<b>64.88</b>	36.61	63.18	35.52
LSTM+Вр	66.53	62.94	35.03	60.93	35.02
LSTM+Эм+Пн	65.89	63.02	36.31	61.33	36.57
LSTM+Эм+Пн+Чр	67.31	64.18	<b>37.64</b>	<b>63.52</b>	<b>37.03</b>
LSTM+Эм+Пн+Чр+Вр	67.10	63.74	37.32	63.21	36.81
<b>Рестораны</b>					
LSTM	62.19	61.64	31.96	83.72	35.84
LSTM+Эм	62.43	61.37	32.74	82.55	36.73
LSTM+Пн	63.01	62.09	32.68	83.23	36.20
LSTM+Чр	64.13	62.48	33.99	<b>85.02</b>	37.12
LSTM+Вр	62.25	61.33	31.83	82.80	35.91
LSTM+Эм+Пн	62.44	61.53	33.20	82.14	36.90
LSTM+Эм+Пн+Чр	<b>64.89</b>	<b>63.12</b>	<b>35.31</b>	84.09	37.53
LSTM+Эм+Пн+Чр+Вр	64.81	62.90	35.15	83.56	<b>37.94</b>

В таблице использованы следующие сокращения:

- «Эм» – символы эмоций;
- «Пн» – пунктуация;
- «Чр» – части речи;
- «Вр» – время глаголов;

Можно сформулировать общий вывод, что признак принадлежности к частям речи оказывает явное позитивное влияние на точность модели – средний прирост F1–меры составляет 1.35%.

Применение морфологического признака времени для глагола не обладает заметным эффектом.

Логично подтверждается гипотеза о том, что обозначения эмоций в тексте несут в себе ценную информацию для задач определения тональности. Использование данного признака позволяет почти на 1% увеличить значение F1–меры. Менее информативна пунктуация в тексте. Вероятно, из-за того, что нельзя однозначно сопоставить знак пунктуации классу тональности, как это практически наверняка можно сделать с эмоциями. Для задачи извлечения аспектных терминов, однако, учет этих признаков не улучшает модель.

Обобщая сказанное, можно утверждать, что признак части речи имеет позитивный эффект для всех задач. Для определения тональности полезным также является учет обозначений эмоций в тексте, и, небольшого улучшения качества можно добиться за счет использования пунктуации.

### 3.6. Рекомендации для дальнейших исследований

Эксперименты показали, что подход использования даже меньшего по размеру специализированного корпуса текстов для получения векторов слов обладает неоспоримым преимуществом. В целях данной работы было использовано лишь несколько наиболее распространенных интернет-ресурсов для создания коллекции отзывов. Одно из возможных направлений развития исследования – получение большого корпуса отзывов на русском языке, сопоставимого по размеру с новостной базой текстов.

Для целей получения качественной модели оправдан подход использования инженерно-лингвистических правил совместно с методами глубокого обучения, что уже было продемонстрировано в работе [32]. Однако, стоит понимать, что это решение фундаментально подрывает одно из главных преимуществ использования подходов глубокого обучения – универсальность. Лингвистические правила обычно привязаны к конкретному языку или предметной области.

Использование в модели дополнительной информации, например, деревьев разбора и морфологических признаков, может обладать позитивным эффектом, что можно увидеть на примере частей речи.

Кроме всего перечисленного, также имеет смысл проведение дальнейших экспериментов с тюнингом алгоритмов – инициализацией начальных весов, регуляризацией, процедурой оптимизации и т.д.

## Заключение

В данной работе рассматривались набирающие в последнее время популярность алгоритмы глубокого обучения применительно к задаче аспектного анализа отзывов с использованием данных конкурса SentiRuEval.

По результатам экспериментов извлечения аспектных терминов было установлено, некоторые модели нейронных сетей способны существенно превзойти по качеству часто применяющийся на практике метод CRF. Нейронные сети при этом обладают гибкостью в использовании и не требуют генерации многочисленных признаков вручную.

Однако, в задаче определения тональности аспектного термина или аспектной группы методы глубокого обучения значительно уступили другим более традиционным классификаторам – SVM и Gradient Boosting Classifier. Особенно большой разрыв заметен при сравнении со сверточными нейронными сетями.

Наиболее оптимальной архитектурой с точки зрения F1–меры почти во всех случаях оказалась LSTM рекуррентная нейронная сеть, иногда – GRU. При этом для задачи поиска аспектов более эффективной стала конфигурация с одним слоем скрытого состояния и прямыми связями. Для задачи определения тональности было предпочтительней использовать рекуррентные нейронные сети с дополнительным слоем с обратными связями.

Эксперименты подтвердили гипотезу о том, что применение специализированной коллекции текстов по заданной тематике в значительной степени положительно влияет на качество работы алгоритма. Кроме того, использование признаков о принадлежности к частям речи обладает существенным положительным эффектом в решении всех задач конкурса. Дополнительный учет символов эмоций также может быть полезен для решения определения эмоциональной окраски в отзывах.



## Список литературы

1. Arkhipenko K. et al. Comparison of neural network architectures for sentiment analysis of Russian tweets. Proc. of the International Conference “Dialogue 2016”, 2016.
2. Bagheri A., Saraee M., de Jong F. An unsupervised aspect detection model for sentiment analysis of reviews. Natural Language Processing and Information Systems. Springer: Berlin Heidelberg, 2013, p. 140-151.
3. Cho K., Merrienboer B., Bahdanau D. On the properties of neural machine translation: Encoder-decoder approaches, 2014.
4. Choi K. et al. Joint extraction of entities and relations for opinion recognition. Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2006), 2006.
5. Frantzi K. et al. Automatic recognition of multi-word terms: the C-value/NC-value method. International Journal on Digital Libraries, 3(2), 2000, p. 115-130.
6. Gal Y., Ghahramani Z. A theoretically grounded application of dropout in recurrent neural networks, 2016.
7. Glavaš G., Korencić D., Šnajder J. Aspect-oriented opinion mining from user reviews in Croatian. Proc. of BSNLP workshop, ACL-2013, 2013, p. 18-23.
8. Graves A. Generating sequences with recurrent neural networks, 2013.
9. Hu M., Bing L.. Mining and summarizing customer reviews. Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), 2004.
10. Irsoy O., Cardie C. Opinion mining with deep recurrent neural networks. Proc. of the 2014 Conference on Empirical Methods in NLP, 2014, p. 720-728.
11. Ivanov V. et al. Extracting aspects, sentiment, and categories of aspects in user reviews about restaurants and cars. SentiRuEval, 2015.
12. Jakob, Niklas, I.Gurevych. Extracting opinion targets in a single- and cross-domain setting with conditional random fields. Proc. of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010), 2010.
13. Jebbara S., Cimiano P. Aspect-based sentiment analysis using two-step neural network architecture. Semantic Computing Group, Germany, 2015.
14. Johansson R. et al. Reranking models in fine-grained opinion analysis. Proc. of the International Conference on Computational Linguistics (COLING-2010), 2010.
15. Karpov I. et al. Entity based sentiment analysis using syntax patterns and convolutional neural network. Proc. of the International Conference “Dialogue 2016”, 2016.

16. Kim S., Hovy E. Automatic identification of pro and con reasons in online reviews. Proc. of COLING/ACL 2006 Main Conference Poster Sessions (ACL-2006), 2006.
17. Kiss T., Strunk J. Unsupervised multilingual sentence boundary detection. Association for Computational Linguistics, 2006.
18. Kobayashi N. et al. Extracting aspect evaluation and aspect-of relations in opinion mining. Proc. of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2007.
19. Kovelamudi S., Ramalingam S., Sood A., Varma V. Domain independent model for product attribute extraction from user reviews using Wikipedia. Proc. of the 5th International Joint Conference on Natural Language Processing (IJCNLP-2010), 2011, p. 1408-1412.
20. Ku L. et al. Opinion extraction, summarization and tracking in news and blog corpora. Proc. of AAAI-CAAW'06, 2006.
21. Lakkaraju H., Socher R., Manning C. Aspect specific sentiment analysis using hierarchical deep learning, 2014.
22. Li S. et al. Comparable entity mining from comparative questions. Proc. of Annual Meeting of the Association for Computational Linguistics (ACL-2010), 2010.
23. Liu B., Zhang L. A survey of opinion mining and sentiment analysis. Mining Text Data.: Springer US. 2012, p. 415-463.
24. Liu P., Joty S., Meng H. Fine-grained opinion mining with recurrent neural networks and word embeddings, 2014.
25. Long C. et al. A review selection approach for accurate feature rating estimation. Proc. of Coling 2010: Poster Vol., 2010.
26. Loukachevitch N. et al. SentiRuEval: Testing object-oriented sentiment analysis systems in Russian, 2015.
27. Mayorov V. et al. A high precision method for aspect extraction in Russian. SentiRuEval-2015, 2015.
28. Meduza [Электронный ресурс]. – Электрон. дан. – URL: <https://meduza.io/> (дата обращения: 13.05.2017).
29. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. Proc. of Workshop at ICLR, 2013.
30. Neviarouskaya A., Prendinger H., Ishizuka M. Recognition of affect, judgment, and appreciation in text. Proc. of the 23rd International Conference on Computational Linguistics (COLING-2010), 2010, p. 806-814.

31. Popescu A., Etzioni O.. Extracting product features and opinions from reviews. Proc. of Conference on Empirical Methods in Natural Language Processing (EMNLP-2005), 2005.
32. Poria S., Cambria E., Gelbukh A. Aspect extraction for opinion mining with a deep convolutional neural network, 2016.
33. Qiu G., Liu B., Bu J., Chen C. Opinion word expansion and target extraction through double propagation. Computational Linguistics. V. 1, No 1., 2011, p. 1-18.
34. Riloff E. Automatically constructing a dictionary for information extraction tasks. Processing of AAAI-2003, 2003.
35. Ruppenhofer J., Somasundaran S., Wiebe J. Finding the sources and targets of subjective expressions. Proc. of LREC, 2008.
36. RusVectors [Электронный ресурс]. – Электрон. дан. – URL: <http://rusvectors.org/ru/> - (дата обращения: 20.03.2017).
37. San Vicente I. et al: A modular and flexible ABSA platform. Proc. of the 9th International Workshop on Semantic Evaluation, 2015, p. 748–752.
38. Santos C., Xiang B., Zhou B. Classifying relations by ranking with convolutional neural networks. Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Vol. 1, 2015, p. 626–634.
39. SentiRuEval [Электронный ресурс]. – Электрон. дан. – URL: <http://www.dialog-21.ru/> (дата обращения 12.01.2017).
40. Tarasov D. Deep recurrent neural networks for multiple language aspect-based sentiment analysis of user reviews. SentiRuEval, 2015.
41. Titov I., McDonald R. A joint model of text and aspect ratings for sentiment summarization, 2008.
42. Toh Z., Su J. Improving aspect based sentiment analysis using neural network features. SemEval, 2016.
43. Wang B., Liu M. Deep learning for aspect-based sentiment analysis, 2015.
44. Wikipedia [Электронный ресурс]. – Электрон. дан. – URL: <http://wikipedia.org/> (дата обращения: 11.02.2017).
45. Xu L., Lin J., Wang L., Yin C., Wang J. Deep convolutional neural network based approach for aspect-based sentiment analysis. Advanced Science and Technology Letters. Vol. 143, 2017, p.199-204.
46. Yu J., Zha Z., Wang M., Wang K., Chua T. Domain-assisted product aspect hierarchy generation: towards hierarchical organization of unstructured consumer reviews. Proc. of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011, p. 140-150.

47. Yu J. et al. Domain-assisted product aspect hierarchy generation: towards hierarchical organization of unstructured consumer reviews. Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2011), 2011.
48. Zhai Z., Liu B., Xu H., Jia P. Grouping product features using semi-supervised learning with soft-constraints. Proc. of Coling-2010, 2010, p. 1272-1280.
49. Zhai Z., Liu B., Xu H., Jia P. Clustering product features for opinion mining. Proc. of the fourth ACM International Conference on Web search and data mining. ACM. 2011, pp. 347-354.
50. Zhang L., Liu B., Lim S., O'Brien-Strain E. Extracting and ranking product features in opinion documents. Proc. of International Conference on Computational Linguistics (COLING-2010), 2010, p. 1462-1470.
51. Zhao W., Jing J., Hongfei Y., Li X. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. Proc. of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010), 2010, pp. 56-65.
52. Zhou X., Wan X., Xiao J. Representation learning for aspect category detection in online reviews. Proc. of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.
53. Zhu J. et al. Multi-aspect opinion polling from textual reviews. Proc. of ACM International Conference on Information and Knowledge Management (CIKM-2009), 2009.
54. Zhuang L., Jing F., Zhu X. Movie review mining and summarization. Proc. of ACM International Conference on Information and Knowledge Management (CIKM-2006), 2006, p. 43-50.