Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра алгоритмических языков

Аникевич Юлия Вадимовна

Методы извлечения аспектов из мнений пользователей

Магистерская диссертация

Научный руководитель:
к.ф.-м.н.
Н.Э. Ефремова

Москва, 2023

Аннотация

В рамках магистерской диссертации рассмотрена задача извлечения аспектов из мнений пользователей. Для ее решения программно реализованы четыре метода извлечения аспектов, являющиеся методам обучения без учителя. Проведено экспериментальное исследование их работы на наборах мнений о ресторанах и мобильных телефонах. Сделаны выводы о применимости выбранных методов для решения поставленной задачи.

Результаты данной диссертации могут быть полезны для исследователей, занимающихся анализом мнений, а также для разработчиков, работающих в области обработки естественного языка.

Содержание

1 Введение

Анализ **мнений** пользователей, т.е. их суждений по поводу некоторого объекта, услуги, продукта, товара, организации и пр. является актуальной задачей, направленной на масштабную обработку мнений, собранных из различных источников. Как правило, цель такой обработки – улучшение и распространение какого-либо продукта или услуги [1]. На основе пользовательских мнений можно выявлять актуальные запросы и изучать пожелания клиентов, искать способы повышения их лояльности, эффективнее проводить маркетинговые кампании.

Сейчас для того, чтобы оставить письменный отзыв или дать обратную связь, существуют десятки специальных платформ: Zoon, Яндекс и Яндекс.Бизнес, 2Gis, Flamp, Otzovik, IReccomend и многие другие. Это показывает высокую востребованность и интерес к отзывам и рекомендациям не только от лица компаний, а также и от лица потребителей и клиентов.

Учитывая сложившуюся ситуацию, текущий объем потребительских отзывов и количество платформ, на которых они могут быть представлены, их качественная обработка вручную сейчас невозможна. Следовательно, запрос на разработку автоматического решения, которое позволит оперативно извлекать и проводить анализ мнений, скрытых за слабоструктурированными текстами, не теряет своей актуальности и стимулирует развитие методов решения задачи анализа мнений [2].

Одним из примеров анализа мнений является анализ тональности, в котором выделяются два подхода: объектно-ориентированный и аспектно-ориентированный [3]. К первому относится определение отношения автора мнения к обсуждаемой в тексте **сущности** (объекту, продукту, сервису и т.п.) в целом. Ко второму – определение отношения автора к характеристикам анализируемой сущности.

Для примера рассмотрим следующий фрагмент отзыва: *«Суп был очень вкусный. Испортил настроение некомпетентный официант»*.

В предложенном фрагменте в качестве сущности выступает некоторый *ресторан* и рассматриваются две его характеристики: *еда* и *персонал*. О *еде* посетитель выразил положительное мнение, о *персонале* высказался негативно. Это приводит нас к идее о том, что иногда мнение о каждой из характеристик следует рассматривать отдельно друг от друга, поскольку они могут существенно различаться.

Выделенные в примере выше характеристики (*еда* и *персонал*) принято называть **аспектами**, или **аспектными категориями**. В их роли, как правило, выступают атрибуты сущности, т.е. качества и свойства, которые ее характеризуют. Например, для сущности *отель* оцениваемыми аспектами обычно являются: *сервис*, *интерьер*, *расположение*, *цена* и т.д. Также, часто в текстах можно встретить оценку сущности в целом, как, например во фразе *«прекрасный ресторан»*.

В свою очередь, слова и выражения, которые обозначают аспекты в тексте, называют **аспектными терминами**. Например, для аспекта *еда* аспектными терминами могут являться: *«аромат»*, *«вкус»*, *«порция»*, *«суп»* и т.п. На Рисунке 1 приведен пример анализа предложения *«Салат был вкусный»*. Система смогла найти в предложении аспектный термин *«салат»* и тональное слово *«вкусный»* и сопоставить их с заранее определенным аспектом *еда* и полярностью настроения *POS* (позитивный) соответственно.

Таким образом, в аспектно-ориентированном подходе к решению задачи анализа тональности предполагается, что текст включает в себя оценку одной сущности, но сущность рассматривается на уровне различных аспектов.

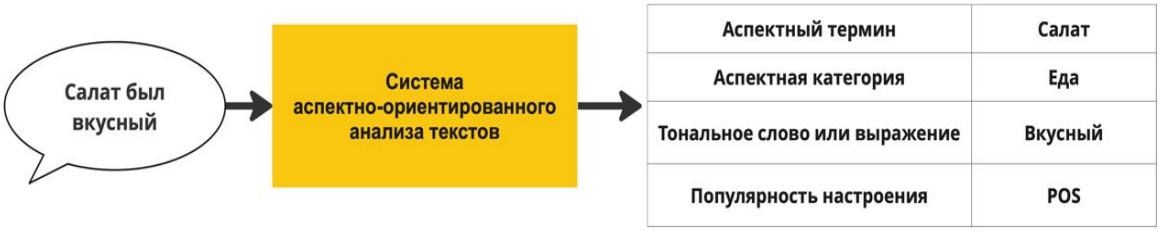

Рис. 1: Пример анализа тональности предложения по аспектам

Впервые задача аспекто-ориентированного анализа тональности текстов была представлена в рамках четвертого международного семинара SemEval [4] в 2014 году. В нашей стране исследования подобного характера были проведены в 2015 году в рамках соревнований SentiRuEval [5] по тестированию систем анализа тональности текстов на русском языке по отношению к заданной сущности.

В ходе данных соревнований были рассмотрены возможности проведения анализа для двух сущностей: *ресторанов* и *автомобилей*. Основная цель участников состояла в том, чтобы предложить методы, позволяющие найти слова и выражения, обозначающие аспектные термины, и определить их тональность и аспектную категорию.

В рамках данной магистерской диссертации рашалась задача извлечения аспектов из мнений пользователей. Для рассмотрения были выбраны методы, которые могут быть применины к текстам отзывов из любой предметной области, т.е. о любой сущности.

По аналогии с подходом, предложенным в рамках соревнований SentiRuEval, решаемая задача была разбита на следующие подзадачи:

1. Выделение аспектных терминов.
2. Формирование аспектных категорий.
3. Маркировка предложений аспектами.

2 Постановка задачи

В данной магистерской диссертации была рассмотрена задача извлечения аспектов из мнений пользователей. Для ее решения было необходимо:

1. Провести обзор современных методов решения задачи извлечения аспектов.
2. По результатам обзора выбрать методы извлечения аспектов из мнений на русском языке и программно их реализовать.
3. Подготовить наборы мнений о двух различных сущностях.
4. Провести экспериментальное исследование качества работы реализованных методов и проанализировать полученные результаты.

3 Обзор методов извлечения аспектов из мнений пользователей

Напоминим, что в рамках данной магистерской диссертации рассматривалась задача извлечения аспектов без привязки к конкретной сущности. Решение задачи подразумевало выделение аспектных терминов, формирование аспектных категорий, а также маркировку предложений полученными аспектными категориями. Вообще, указанные подзадачи могут решаться как вместе, так и по-отдельности. В данной главе рассмотрены современные методы их решения.

Сейчас выделяют следующие основные подходы к извлечению аспектов из текстов на естественном языке [6]:

1. Инженерный подход.
2. Статистический подход.
3. Машинное обучение.

Методы инженерного подхода используют для извлечения аспектных терминов и аспектов заранее определенные шаблоны и правила, которые могут быть созданы вручную экспертами или сгенерированы автоматически на основе языковых и структурных характеристик текста. Например, аспектными терминами могут считаться последовательности существительных (*«заряд батареи»*), пары прилагательное и существительное (*«открытый бассейн»*), последовательности существительных и предлогов (*«блюда из овощей»*) и т.д. [7].

В работе [8] рассмотрены методы формирования аспектных категорий, опирающиеся на семантические отношения между словами. Первый реализованный метод формирует аспекты, используя только отношения синонимии (т.е. аспектные термины, которые находятся в отношениях синонимии, группируются вместе). Во втором методе к одному аспекту относятся слова, состоящие в отношении синонимии и отношении *is-a*. Третий метод использует отношение синонимии, *is-a* и *part-of*.

При правильной разработке правил методы инженерного подхода могут давать высокое качество извлечения аспектов и аспектных терминов, особенно когда речь идет

о какой-то конкретной сущности. В то же время, для их работы необходимо постараться заранее учесть все возможные варианты выражения в тексте аспектов, которые зависят от контекста, стиля написания и т.д., что может быть достаточно трудоемким процессом. Кроме того, среди выделяемых таким образом терминов обычно оказывается много слов и словосочетанияй, ими не являющихся. Поэтому методы инженерного подхода чаще всего комбинируются с методами других подходов.

3.1 Методы статистического подхода

Методы, основанные на статистике, рассматривают задачу извлечения аспектных терминов как задачу извлечения статистически значимыми для анализируемых мнений слов и словосочетаний. По данным [9] 60-70% аспектных терминов являются именными группами, поэтому методы данного подхода в первую очередь считают терминами существительные и именные словосочетания. Для достижения наилучшего качества извлечения аспектов методы, основанные на статистике, обычно используются в комбинации с другими методами.

Как правило, статистический метод содержит следующие шаги [10]:

1. Извлечение кандидатов в аспектные термины.
2. Вычисление для них статистической характеристики.
3. Отбор кандидатов, значение статистической характеристики у которых выше заданного порога.

Первые попытки решения задачи извлечения аспектов с помощью статистических методов были основаны на классическом подходе к решению задачи извлечения информации, заключающемся в выявлении в тексте часто встречающихся слов. К примеру, в работе [9] использовался метод, общая идея которого сводится к поиску существительных либо именных словосочетаний и выбору в качестве аспектов наиболее частотных из них. Однако, очевидно, что такой метод упускает низкочастотные аспектые термины.

Для того чтобы извлекать и низкочастотные термины была предложена статистическая мера TF-IDF [11]:

$$TF - IDF(t, d, D) = TF(t, d) * IDF(t, D)$$

Здесь TF (Term Frequency) вычисляется как частота встречаемости термина в документе, где $count(t, d)$ – количество употреблений термина t в документе d, а T_d – множество терминов документа d:

$$TF(t, d) = \frac{count(t, d)}{\sum_{i \in T_d} count(i, d)}$$

В свою очередь IDF вводится так, чтобы редко встречающиеся в документах термины имели больший вес:

$$IDF(t, D) = log \frac{|D|}{|\{d \in D | t \in T_d\}|},$$

где |D| – общее число документов в корпусе, $|\{d \in D | t \in T_d\}|$ – число документов в корпусе, в которых встречается термин t.

В работе [12] сравнивается частота употребления кандидатов в аспектные термины в основном и контрастном корпусе. В основе предлагаемого метода лежит гипотеза о том, что частота употребления аспектных терминов в основном корпусе должна быть выше, чем в контрастном.

Статистический метод, предложенный в работе [13], опирается на значение p-support. Значение p-support для термина t – это количество предложений, содержащих t (например, *«батарея»*), исключая предложения, содержащие другой аспектный термин t', который включает t в свой состав (например, *«заряд батареи»*). После выделелеления из текста всех существительных и именных словосочетаний из них формируются новые кандидаты: в рамках одного предложения рассматриваются пары и триплеты выделенных слов и словосочетаний. Далее вычисляется значение p-support для каждого нового кандидата, и те кандидаты, значение p-support у которых ниже некоторого заданного порога, удаляются из дальнейшего рассмотрения.

К достоинствам методов, основанных на статистике, следует отнести хорошую масштабируемость, отсутствие потребности в большом количестве размеченных данных, а также стоимость разработки и внедрения статистических методов, которая может быть существенно ниже, чем для методов других подходов, поскольку не требуют привлечения большого количества экспертов.

В качестве недостатков статистических методов можно обозначить пропуск низкочастотных терминов и трудность обработки сложных контекстов.

3.2 Методы машинного обучения

Для решения задачи извлечения аспектов из текста также часто применяют методы машинного обучения [6]. Существуют несколько наиболее популярных методов, решающих поставленную задачу либо как бинарную задачу классификации, либо как задачу классификации последовательностей (англ. sequential classification). В последние годы при решении рассматриваемой задачи активно используются нейронные сети.

При постановке задачи извлечения мнений как задачи классификации требуется классифицировать извлеченное из текста слово или словосочетание как аспектное или неаспектное. Например, для русскоязычных текстов такой подход был опробован в работе [14]. В ней были рассмотрены такие популярные классификаторы, как метод опорных векторов, наивный байесовский классификатор (бернуллевский и мультиномиальный), логистическая регрессия, дерево решений и случайный лес. Авторами было отмечено, что лучшие результаты показали метод опорных векторов и наивный байесовский классификатор.

Заметим, что специфика решаемой задачи подразумевает необходимость учитывать взаимосвязи между словами, что достаточно сложно при использовании традиционных методов классификации. По этой причине для задачи извлечения мнений часто применяются методы сегментации и разметки последовательностей. К этой группе методов можно отнести метод на основе скрытых марковских моделей, а также метод условных случайный полей, которые, используя входную последовательность терминов, моделирует условное вероятностное распределение для последовательной разметки, при которой аспекты размечаются в корпусе [15, 16].

Однако, из-за необходимости наличия для обучения методов специально размеченных данных эти методы достаточно сложно перенастраивать на различные предметные области. Также, для языков со свободным порядком слов в предложении эти методы будут показывать достаточно скромные результаты.

Последнее время для извлечения аспектов активно используются различные архитектуры нейронных сетей, в том числе глубокие нейронные сети [17], рекуррентные нейронные сети [18] и сверточные нейронные сети [19]. Например, в работе [20] предложена модель, комбинирующая для извлечения аспектов сверточную и рекуррентную нейронные сети.

В [21] работе предложен метод извлечения аспектов без учителя, который использу-

ет векторное представление слов с учетом контекста, а также механизм внимания для того, чтобы снизить вклад незначимых слов во время обучения. Результатом обучения является матрица, в которой каждая строка интерпретируется как вектор аспекта в заданном векторном пространстве. Ближайшие слова к этим векторам являются аспектными терминами, отражающими определенные аспекты.

Еще одним методом извлечения аспектных терминов без учителя является метод на основе контрастивного, или контрасного внимания [22]. Стоит заметить, что для работы метода аспектные категории (точнее, называющие их слова) необходимо определить заранее.

Работу предложенного метода можно разбить на четыре шага:

1. Обработка текстов мнений и получения векторых представлений слов.

2. Извлечение аспектных терминов: рассматриваются наиболее частотные существительные.

3. Использование механизма контрастивного внимания и формирование взвешенной суммы слов предложения.

4. Определение аспектной категории для каждого аспектного термина: вычисляется косинусное сходство для векторных представлений слов.

Кроме того, извлечение аспектов может выполняться на основе статистических тематических моделей, которые основываются на предположении, что каждый текст состоит из набора скрытых тем, а каждая скрытая тема представляет собой вероятностное распределение слов. Результатом применения моделей является набор тем (для нашей задачи – аспектов), представляющих собой список слов (кандидатов в аспектные термины) с вероятностями их отнесения к данной теме (аспекту).

В частности, к методам тематического моделирования относится метод скрытого размещения Дирихле (англ. Latent Dirichlet Allocation, LDA). Пример применения известной базовой модели на основе LDA приведен в работе [23]. Авторы использовали глобальную модель для извлечения именований сущностей, а для извлечения аспектных терминов использовали скользящее окно из слов или предложений. Оказалось, что метод LDA хорошо подходит для определения самих сущностей, но не является эффективным для извлечения аспектов.

Метод тематического моделирования битермов (BiTERM) – это еще одна вероятностная модель тематического моделирования [24]. В отличие от LDA, в котором с определенной темой ассоциируется каждое слово в документе, в методе BiTERM с темой ассоциируются **битермы** – неупорядоченные пары слов, встречающиеся в одном предложении. Это позволяет учитывать больше информации из контекста и улучшает результаты моделирования на коротких текстах, к которым относятся и мнения пользователей.

Дополнительно рассмотрим несколько методов, сочетающих в себе идеи из разных подходов. В работе [25] предложен метод решения задачи формирования аспектных категорий, основанный на использовании семантической близости между словами и векторного представления слов. Подзадача извлечения аспектных терминов решается путем автоматического расширения предопределенного набора терминов для каждого аспекта.

В данном случае список аспектных категорий и базовый набор терминов заранее известны, но поставленную задачу можно решить и без начальной информации. Например, в работе [26] также используется векторное представление слов и проводится кластеризация слов и словосочетаний, являющихся кандидатами в аспектные термины, на базе сходства их контекстов. Для кластеризации используется алгоритм K-means, количество кластеров определяется с помощью метод локтя [27].

В данной главе были рассмотрены основные методы решения задачи извлечения аспектов из мнений пользователей. Методы инженерного подхода, как правило, являются плохо масштабируемыми и выделяют много нерелевантных терминов. Статистические методы могут пропускать низкочастотные термины, что может снижать точность извлечения аспектов, а также давать некорректное определение аспекта в сложных контекстах. Методы машинного обучения с учителем показывают хорошие результаты, но для их работы требуется набор размеченных текстовых данных соответствующей предметной области, что является существенным недостатком.

Поскольку в рамках данной диссертации ставилась задача извлечения аспектов без привязки к конкретной предметной области и рассматриваемой в ней сущности, было решено рассматривать современные методы, не требующие для своей работы корпуса размеченных мнений, а именно:

- Метод, являющийся комбинацией методов, описанных в [9] и [26]. Назовем его FREQ (от англ. frequency – частота).
- Метод тематического моделирования битермов (BiTERM), предложенный в [24] .
- Метод на основе механизма внимания, описанный в [21]. Назовем его ATT (от англ. attention – внимание).
- Метод на основе контрастивного внимания, предложенный в [22]. Назовем его CATT (от англ. contrastive attention — контрастивное внимание).

4 Методы извлечения аспектов и их программная реализация

В данной главе подробно описаны методы, выбранные для решения поставленной в магистерской диссертации задачи извлечения аспектов из мнений пользователей. Для каждого метода указано, каким образом он:

1. Выделяет аспектные термины.
2. Формирует аспектные категорий.
3. Маркирует предложения аспектами.

В конце главы приведено описание программной реализации выбранных методов.

4.1 Метод FREQ

Данный метод извлекает аспектные термины, основываясь на частоте встречаемости слов в текстах, и формирует аспектные категории с помощью алгоритма кластеризации.

На вход методу FREQ задается порог отсечения кандидатов в аспектные термины по частоте, а также требуемое количество кластеров. На выходе метода – кластеры аспектов с соответствующими списками аспектными терминов. Схематично работа метода представлена на Рисунке 2.

Для выделения аспектных терминов из текстов извлекаются все существительные – кандидаты в аспектные термины и подсчитывается частота их встречаемости. Далее отсекаются все кандидаты, у которых частота встречаемости ниже заданного порога.

Для применения алгоритма кластеризации необходимо сначала преобразовать слова в векторные представления, так как алгоритмы кластеризации работают с числовыми признаками. Если данные представлены в виде числовых векторов, то алгоритмы кластеризации могут определять сходство объектов, сравнивая значения их признаков, и создавать кластеры на основе этого сходства.

На основе векторных представлений аспектных терминов формируются аспектные категории, для чего используется алгоритм кластеризации. Благодаря использованию векторного представления в кластеры объединяются семантически близкие аспектные

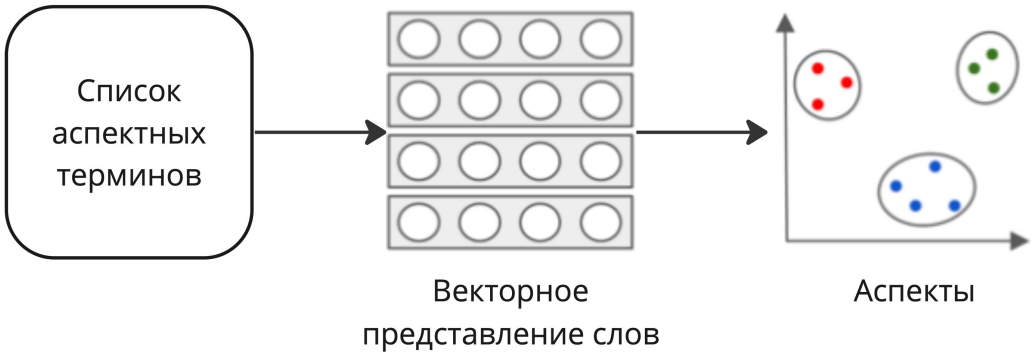

Рис. 2: Схема работы метода FREQ

термины. В итоге, аспектные категории идентифицируются с центроидами найденных кластеров, а каждый кластер содержит список аспектных терминов.

Предложение маркируется тем аспектом, к которому относятся выделенные из него аспектные термины.

4.2 Метод BiTERM

Метод тематического моделирования битермов является моделью тематического анализа. Битерм определяется как неупорядоченная пара слов, встречающихся в одном контексте. Например, если слова *«еда»*, *«суп»* и *«вареники»* часто встречаются друг с другом в одном и том же контексте, можно говорить, что они принадлежат к одному и тому же аспекту. В данной диссертации каждое предложение рассматривается как отдельный контекстный блок, что означает, что любые два слова в одном предложении образуют битерм. Например, битермы предложения *"Модель телефона устарела"* будут выглядеть следующим образом:

(модель, телефон), (телефон, устареть), (модель, устареть)

На вход метода BiTERM подается количество аспектных категорий, параметры распределения Дирихле и количество итераций метода.

После извлечения различных битермов из каждого предложения корпус превращается в набор битермов B. Генеративный процесс в BiTERM в терминах задачи извле-

чения аспектов можно описать следующим образом:

1. Выбрать распределение аспектов $\theta_i \sim Dir(\alpha)$.
2. Для каждого аспекта z:
   - Выбрать распределение слов для аспекта $\phi_z \sim Multinomial(\beta)$.
3. Для каждого битерма b_i из набора битермов B:
   - Выбрать аспект $z \sim Multinomial(\theta)$.
   - Выбрать два слова $w_i, w_j \sim Multinomial(\phi_z)$.

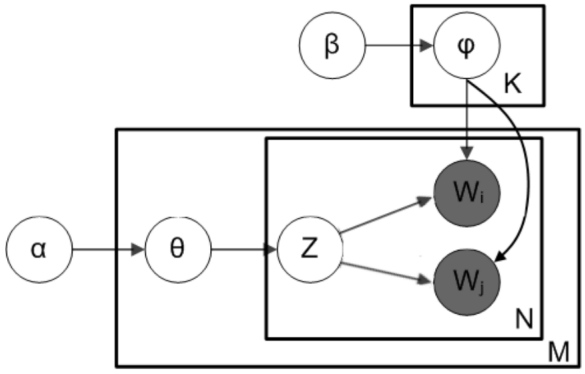

Рис. 3: Графическое представление модели BiTERM

Для определения аспектной категории для каждого предложения, предполагается, что пропорции аспектов в предложении равны математическому ожиданию пропорций битермов, сгенерированных из данного предложения:

$$P(z|d) = \sum_b P(z|d)P(b|d)$$

В свою очередь $P(z|b)$ можно рассчитать по формуле Байеса по параметрам, оцениваемым в методе BiTERM следующим образом:

$$P(z|b) = \frac{P(z)P(w_i|z)P(w_j|z)}{\sum_z P(z)P(w_i|z)P(w_j|z)}$$

Для нахождения $P(b|d)$ используется эмпирическое распределение битермов в предложении:

$$P(b|d) = \frac{n_d(b)}{\sum_b n_d(b)},$$

где $n_d(b)$ – частота битерма b в предложении d.

4.3 Метод ATT

Данный метод предполагает использование автоэнкодера и механизма внимания, который показывает, какие именно слова в тексте при извлечении аспектов имеют наибольшее значение. На вход метода ATT передается количество аспектных категорий K. На выходе получаем матрицу аспектов, каждый строка которой интерпретируется как вектор аспекта, а также список ближайших (в векторном смысле) аспектных терминов для каждого аспекта.

Обучение нейронной сети можно разбить на следующие шаги (см. Рисунок 4):

1. Получение вектора предложения z_s как взвешенной суммы векторов слов.
2. Получение из z_s вектора p_t размерности K – вектора вероятностей отнесения предложения к аспектам.
3. Реконструкция вектора предложения r_s.

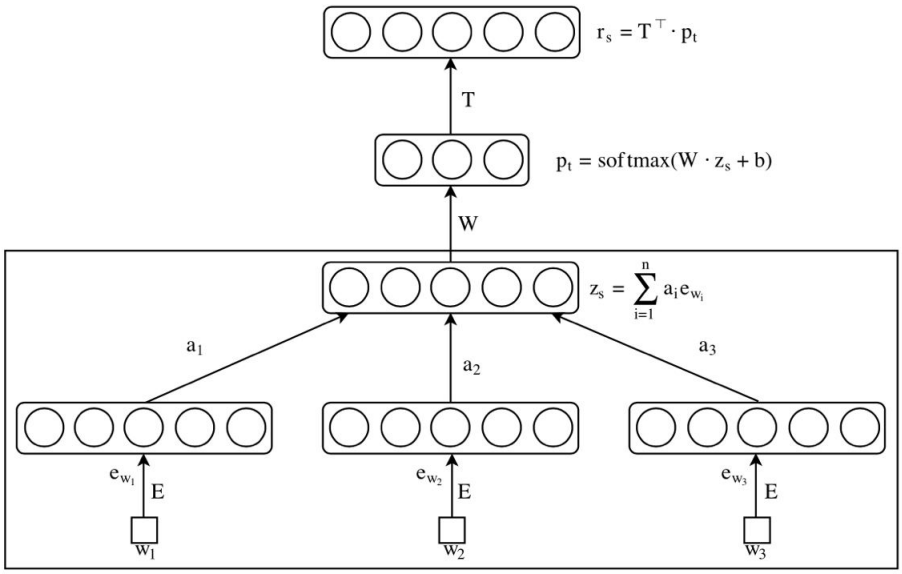

Рис. 4: Схема обучения метода ATT

На первом шаге для каждого слова w получаем его векторное представление $e_w \in \mathbb{R}^d$,

где d – размер векторного пространства, и формируем вектор предложения z_s:

$$z_s = \sum_{i=1}^{n} a_i e_{w_i}$$

Положительный вес a_i для каждого слова w_i в предложении интерпретируется как вероятность того, что w_i является словом, отражающим смысл предложения, и вычисляется согласно следующей формуле:

$$a_i = \frac{exp(d_i)}{\sum_{j=1}^{n} exp(d_j)}$$

$$d_i = e_{w_i}^T \cdot M \cdot y_s,$$

где матрица $M \in \mathbb{R}^{dxd}$ – отображение между y_s и векторным представлением e_{w_i}, которое является параметром обучения, а y_s – среднее арифметическое векторов слов предложения.

На следующем шаге к вектору z_s применяется линейный слой и функция softmax:

$$p_t = softmax(W \cdot z_s + b),$$

где W и b – параметры обучения. На выходе получаем вектор p_t размерности K. Как уже было сказано, данный вектор интерпретируется как вектора вероятностей отнесения предложения к аспектам.

На последнем шаге происходит реконструкция векторного представления предложения путем линейной комбинации векторов из матрицы представлений аспектов $T \in \mathbb{R}^{Kxd}$

$$r_s = T^T \cdot p_t$$

Общая цель обучения – свести к минимуму потери при реконструкции, т.е. разницу между r_s и z_s.

В результате обучения получаем матрицу представления аспектов, где каждый вектор может интерпретироваться как вектор аспекта. Список аспектных терминов может быть получен как список ближайших слов к векторам аспектов в заданном векторном пространстве. Маркировка предложений производится исходя из вектора вероятностей отнесения предложения к аспектам p_t – выбирается наиболее вероятный аспект.

4.4 Метод CATT

В данной диссертации метод на основе контрастивного внимания рассматривается только для решения подзадачи маркировки предложения аспектами. На вход данному методу подаются аспектные категории (точнее, слова, называющие их), порог отсечения кандидатов в аспектные термины и коэффициент масштабирования γ. На Рисунке 5 изображена схема работы метода.

Начальным шагом работы метода является формирование списка аспектных терминов A. В [22] было проведено исследование, показавшее, что хорошими кандидатами в аспектные термины для данного метода являются наиболее частотные существительные. Исходя из этого в данной работе аспектными терминами также были выбраны наиболее частотные существительные.

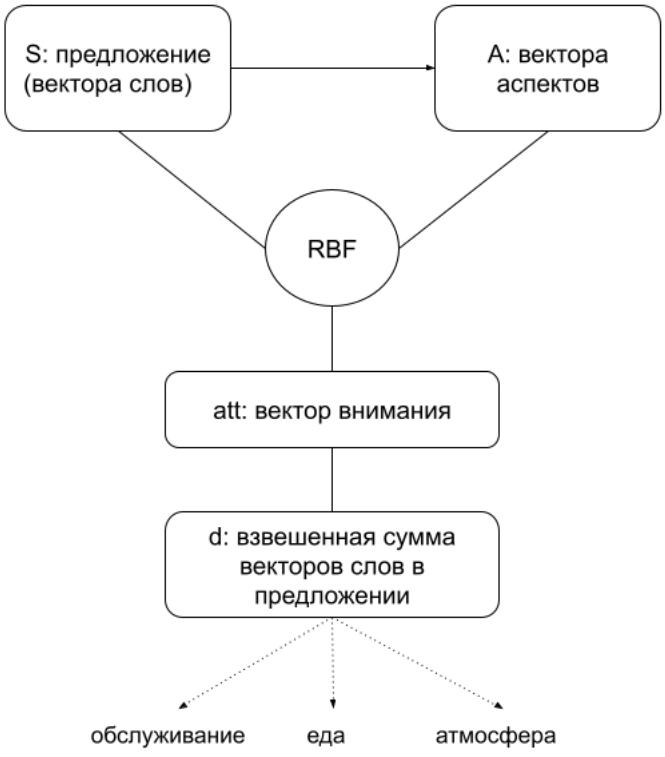

Рис. 5: Схема работы метода CATT

На следующем шаге вычисляется взвешенная сумма векторов слов в предложении, где для расчета коэффициентов используется механизм внимания, позволяющий полу-

чить распределение вероятности отнесения слов в предложении к заданным аспектам. В данном методе используется контрастивное внимание, которое позволяет акцентироваться на словах, которые обозначают наличие аспекта в тексте, независимо от того, к какому именно аспекту они относятся.

Вектор контрастивного внимания может быть вычислен следующим образом:

$$att = \frac{\sum_{a \in A} rbf(w, a, \gamma)}{\sum_{w \in S} \sum_{a \in A} rbf(w, a, \gamma)},$$

где A – список векторов аспектных терминов, S – предложение в виде списка векторов входящих в него слов. Таким образом слова, которые в среднем более похожи на аспекты, получают более высокое значение внимания.

В приведенной выше формуле используется ядро радиальной базисной функции (RBF), которое определяется следующим образом:

$$rbf(x, y, z) = exp(-\gamma ||x - y||_2^2),$$

где x, y – вектора, γ – коэффициент масштабирования.

Важным свойством ядра RBF является то, что оно превращает произвольное неограниченное расстояние, в данном случае квадратичное евклидово расстояние, в ограниченное подобие. Например, независимо от γ, если расстояние между векторами x и y равно 0, значение ядра будет равно 1. По мере увеличения расстояния между векторами, их сходство уменьшается и в конечном итоге асимптотирует к 0. Скорость сходимости зависит от коэффициента масштабирования γ.

На последнем шаге происходит маркировка предложений аспектами. Для этого используется косинусное сходство между вектором предложения и векторами аспектов.

4.5 Программная реализация методов извлечения аспектов

Выбранные методы извлечения аспектных категорий были программно реализованы. В качестве языка программирования был выбран язык Python 3, поскольку он обладает обширным набором библиотек, предназначенных для обработки естественного языка. На Рисунке 6 показана архитектура созданного программного модуля.

На вход программному модулю поступает коллекция отзывов в виде набора текстов. Предобработка каждого текста состоит из следующих этапов:

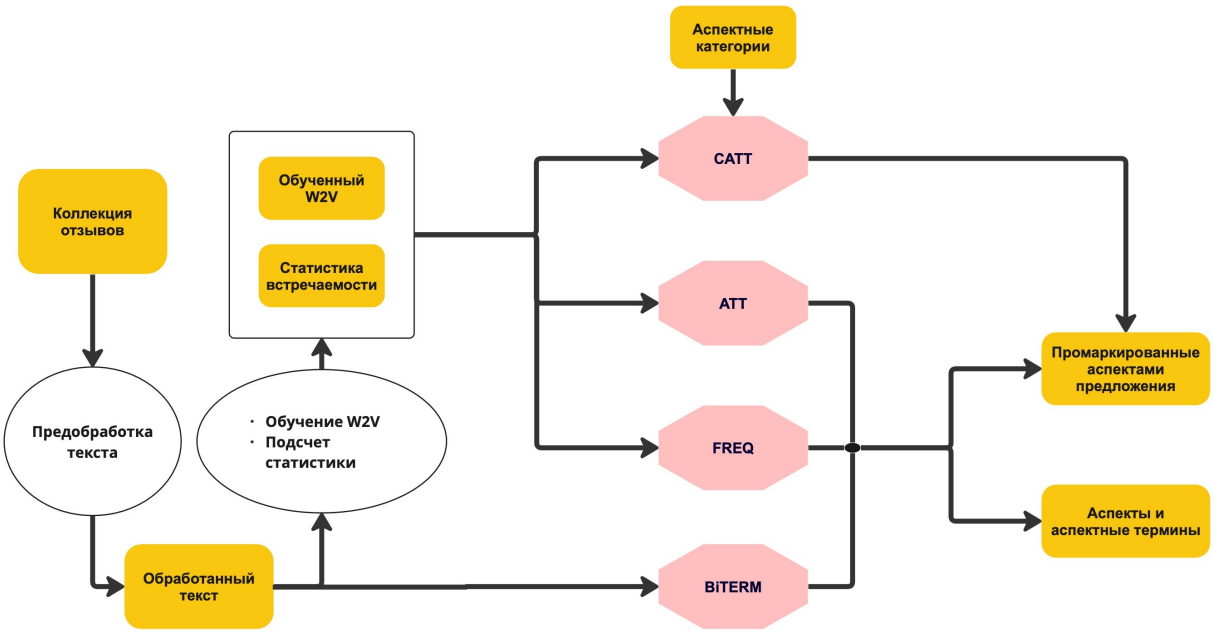

Рис. 6: Архитектура разработанного программного модуля

- Сегментация предложений.
- Удаление пунктуации.
- Приведение слов к нижнему регистру.
- Удаление стоп-слов.
- Лемматизация слов.

Для сегментации текста на предложения, удалении пунктуации и стоп-слов используются возможности библиотеки NLTK [28]. Для частеречной разметки и лемматизации используется библиотека pymorphy2 [29].

В данной работе для получения векторного представления кандидатов в аспектные термины было принято решение использовать дистрибутивно-семантическую модель word2vec [30]. Модель word2vec позволяет представлять слова в виде векторов в n-мерном пространстве, которые обладают следующим свойством: слова, встречающиеся в похожих контекстах, имеют близкие вектора в этом пространстве.

После предобработки входных текстов происходит обучение модели word2vec и подсчет статистики встречаемости существительных в тексте. Результаты сохраняются в

двух файлах: обученная модель сохраняется в .w2v файле, статистика встречаемости существительных – в .json файле.

Полученные файлы поступают на вход реализациям соотвествующих методов, для каждого метода задаются все необходимые параметры. В результате работы методы возвращают два выходных файла: один файл содержит промаркированные аспектами предложения, другой – аспектные термины.

Для представления слов в виде векторов признаков использовалась библиотека gensim [31], для реализации классических методов машинного обучения – библиотека scikit-learn [32], для работы нейронными сетями – библиотека PyTorch [33]. Для кластеризации используется алгоритм K-means [34].

5 Экспериментальное исследование работы методов извлечения аспектов

5.1 Описание наборов данных

Для оценки качества работы предложенных методов извлечения аспектов в данной работе было использовано два набора данных на русском языке, содержащих мнения о ресторанах и мобильных телефонах. Каждый набор был разделен на обучающую (неразмеченную) и тестовую (размеченную) выборки. В тестовой выборке каждому предложению была присвоена метка, указывающая на аспект, выраженный в данном предложении. Например, предложению *"Суши были великолепны"* была присвоена метка *еда*. Аспекты, которые были использованы для разметки, приведены в Таблице 1.

Мобильные телефоны	Рестораны
Камера	Кухня
Аккумулятор	Интерьер
Динамики	Сервис
Цена	Цена
Память	
Дисплей	
Производительность	

Таблица 1: Аспекты, использованные при разметке наборов данных

Наборы мнений о ресторанах были собраны с Google Карт с помощью инструмента Google Maps Reviews Crawler. Набор содержит 419345 мнений и 404 размеченных аспектами предложения. Размеченные предложения взяты из набора, предоставленного в рамках соревнований SentiRuEval 2015 года.

Наборы мнений о мобильных телефонах был взят с платформы Kaggle и содержит 458433 мнений. Разметка тестовой выборки размером 1300 предложений была произведена на платформе Яндекс.Толока.

На Рисунке 6 приведен пример разметки предложений из набора отзывов о ресторанах.

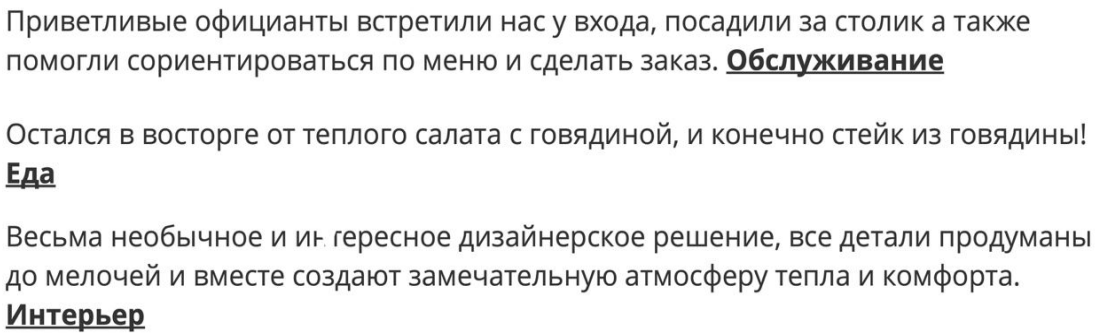

Рис. 7: Пример разметки аспектами ресторанных отзывов

5.2 Меры качества работы методов

5.2.1 Показатель когерентности

Для оценки качества работы методов извлечения аспектов был выбран показатель когеретности – мера, используемая для оценки связности полученных аспектов. Пусть дан аспект z и набор из N самых характерных для него терминов $S^z = \{w_1^z, ..., w_N^z\}$. Тогда показатель когерентности вычисляется следующим образом:

$$C(z; S^z) = \sum_{n=2}^{N} \sum_{l=1}^{n-1} log \frac{D_2(w_n^z, w_l^z) + 1}{D_1(w_l^z)},$$

где $D_1(w)$ – частота встречаемости слова w в предложениях, $D_2(w_1, w_2)$ – частота совместной встречаемости слов w_1 и w_2 в предложениях.

Можно провести аналог между когерентностью и мерой внутрикластерной схожести. Более высокое значение показателя когерентности указывает на лучшую **интерпретируемость** аспекта, т.е. на то, что отнесенные к одному аспекту термины больше семантически связаны друг с другом, чем с терминами из других аспектов.

Также используется средний показатель когерентности, который вычисляется как среднее значение показателей когерентности по всем аспектам:

$$C_K = \frac{1}{K} \sum_{k=1}^{K} C(z_k; S^{z_k})$$

5.2.2 Экспертная оценка

Экспертная оценка является важным этапом оценки качества полученных аспектов. Она позволяет понять, насколько человеку легко интерпретировать результаты, выданные методами.

Экспертам предоставлялся список из n упорядоченных по близости к аспекту терминов. Эксперты определяли название соответствующего аспекта и отмечали, какие из предложенных терминов отражают этот аспект. Пример такой разметки можно увидеть в Приложении А.

Количество отмеченных экспертами терминов обозначим m. Для оценки использовалась мера $precision@n(p@n)$, которая была впервые использована в [35], мера выражается следующей формулой:

$$precision@n = \frac{m}{n}$$

5.2.3 Оценка маркировки предложений

Для оценки качества разметки предложений аспектами использовались точность, полнота и F1-мера.

Точность (*precision*) – это мера того, как много предложений, отнесенных к определенному аспекту, действительно относятся к этому аспекту. Точность показывает долю правильно промаркированных предложений из всех предложений, которые были отнесены к данному аспекту, и выражается формулой:

$$\frac{TP}{TP + FP},$$

где TP (True Positive) – количество правильно промаркированных предложений, FP (False Positive) – количество неправильно промаркированных предложений.

Полнота (*recall*) – это мера того, как много предложений, относящихся к определенному аспекту, были действительно определены в качестве таковых. Полнота показывает долю правильно промаркированных предложений из всех предложений, которые действительно относятся к данному аспекту, и выражается формулой:

$$\frac{TP}{TP + FN},$$

где FN (False Negative) – количество предложений, которые относятся к данному аспекту, но были ошибочно отнесены к другому аспекту.

F1-мера является гармоническим средним между точностью и полнотой. Чем ближе значение F1-меры к 1, тем лучше качество работы метода. F1-мера высчилялась как:

$$f1 = 2\frac{precision \cdot recall}{precision + recall}$$

5.3 Полученные результаты и их анализ

Для эспериментального исследования работы методов на двух наборах данных было выбрано число аспектов равное 10. Еще раз напомним, что подзадачи извлечения аспектных терминов и формирования аспектов решаются методами FREQ, BiTERM и ATT. Подзадача маркировки предложений решается всеми предложенными методами. Результаты, представленные для каждого из методов, являются средним значением результатов за 3 запуска.

Для метода FREQ порог отсечения кандидатов в аспектные термины был выбран равным 10. Параметры, используемые для обучения модели word2vec, являются стандартными и не подвергались специальной настройке.

Для метода BiTERM были выбраны параметры распределения Дирихле $\alpha = 50/\mathrm{K}$ и $\beta = 0.1$. Было произведено 1000 итераций для обучения тематической модели. Выбор параметров распределения Дирихле и числа итераций происходил на основе стандартных рекомендаций и экспериментов с разными значениями.

Для метода ATT матрица векторных представлений аспектов была инициализирована центроидами кластеров, полученных кластеризацией всех слов заданного векторного пространства с помощью алгоритма K-means.

Для метода CATT был определен коэффициент масштабирования γ, равный 0.1, и пороговое значение отсечения кандидатов в аспектные термины, равное 10.

На Рисунке 8 указано значение показателя когереностности для методов FREQ, BiTERM и ATT. Видно, что метод ATT превосходит два других метода на обоих наборах данных. Метод FREQ, примененный к векторным представлениям слов и разделяющий слова на кластеры с учетом их семантической близости, по результатам сравнения превосходит метод BiTERM.

Можно заметить, что при увеличении числа терминов, отнесенных к аспекту, значение когерентности незначительно уменьшается. Такое явление связано с тем, что с увеличением числа терминов возрастает шум, но тем не менее большая часть слов все

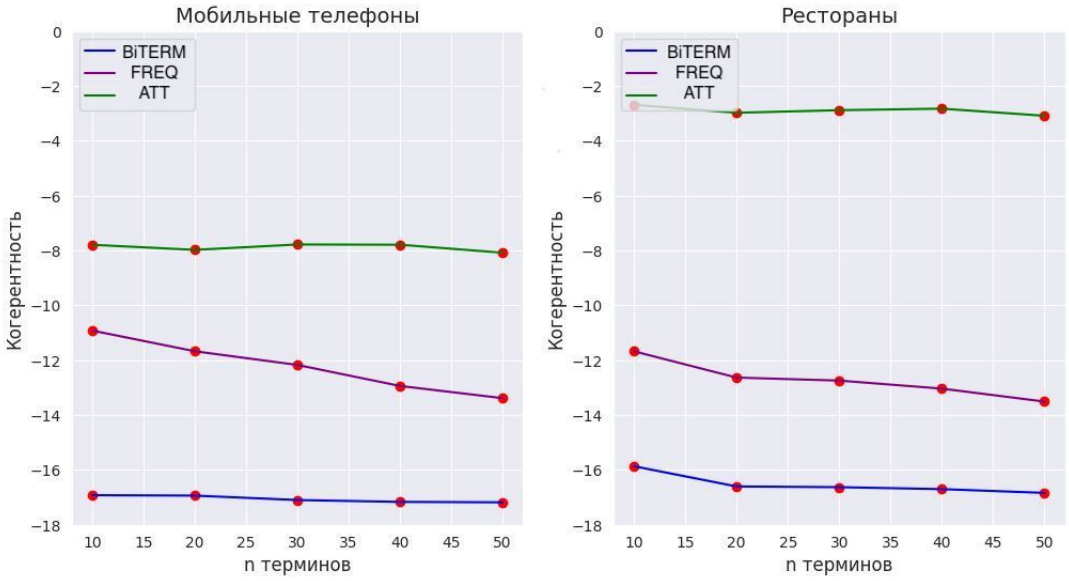

Рис. 8: Среднее значение когерентности в зависимости от числа терминов

еще остается семантически связанной между собой и соответствует определенным аспектам.

Для экспертной оценки качества полученных аспектов было привлечено четыре независимых асессора. Каждый асессор проводил независимую работу, что позволило получить более объективную оценку. Аспект считался верно извлеченным, если его смогли определить более половины асессоров. Результаты интерпретации полученных аспектов для ресторанов и мобильных телефонов приведены в Приложении Б на Рисунке 11 и на Рисунке 12 соответственно.

Таким образом, после экспертной оценки была получена общая картина качества извлеченных аспектов, представленная на Рисунке 8; диаграмма справа показывает результаты работы методов на наборе отзывов о ресторанах, диаграмма слева – результаты работы методов на наборе отзывов о мобильных телефонах. Выяснилось, что часть аспектов была выделена правильно, а часть аспектов ни смог интерпретировать ни один эксперт. Из 10 рассматриваемых в экспериментальном исследовании аспектов экспертами однозначно были интерпретированы:

- для метода FREQ: 7 для ресторанной тематики и 8 для мобильных телефонов;
- для метода BiTERM: 5 для ресторанной тематики и 5 для мобильных телефонов;

- для метода ATT: 9 для ресторанной тематики и 7 для мобильных телефонов.

Видно, что метод ATT превосходит другие методы для ресторанных аспектов, а FREQ показал более хороший результат для мобильных телефонов. Тем не менее, оба метода достигли высоких результатов и могут быть полезны при задаче формирования аспектов.

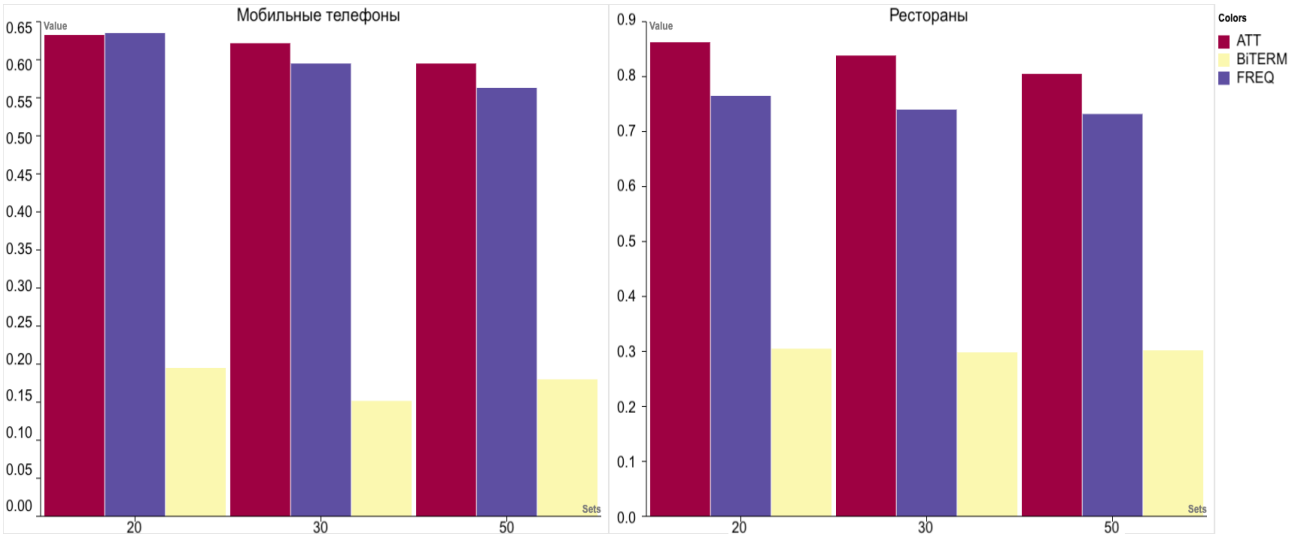

Рис. 9: Значение метрики precision@n в зависимости от числа терминов

Для оценки маркировки предложений аспектами были использованы такие метрики, как precision, recall и F1-мера. Задача маркировки аспектами на уровне предложений производится единственным аспектом для избежания неоднозначности. В тестовой выборке, связанной с ресторанами, присутствуют предложения, отмеченные только одним аспектом, в количестве 74% от общего числа предложений. В то же время, в тестовой выборке, связанной с аспектами мобильных телефонов, присутствуют предложения с только одним аспектом в 81% случаев. Для случаев, когда в тестовой выборке предложения содержат более одного аспекта, корректным будет считаться определение любого из аспектов, указанных в тестовой разметке.

Результаты маркировки предложений аспектами представлены в Таблицах 2 и 3. Данные результаты позволяют увидеть, что ATT показал лучшее значение F1-меры для ресторанных аспектов *Кухня*, *Сервис*, *Цена* и для аспектов мобильных телефонов *Аккумулятор*, *Цена*, *Дисплей*. Метод CATT показал лучшие результаты меры F1 на аспектах *Интерьер*, а также *Камера*, *Динамики*, *Память*. Метод FREQ имеет лучшее

значение для аспекта *Производительность* для мобильных телефонов. Можно сделать вывод, что оба метода ATT и CATT хорошо работают для задачи маркировки предложений аспектами.

Аспект	**Метод**	**Precision**	**Recall**	**F1**
Кухня	CATT	0.848	0.885	0.866
	ATT	**0.988**	0.864	**0.922**
	FREQ	0.855	**0.974**	0.911
	BiTERM	0.584	0.730	0.649
Интерьер	CATT	**0.855**	0.689	**0.763**
	ATT	0.559	**0.988**	0.715
	FREQ	0.772	0.667	0.716
	BiTERM	0.435	0.871	0.580
Сервис	CATT	**0.831**	0.713	0.767
	ATT	0.820	0.783	**0.801**
	FREQ	0.615	**0.889**	0.727
	BiTERM	0.736	0.625	0.676
Цена	CATT	0.634	0.802	0.708
	ATT	**0.962**	0.631	**0.762**
	FREQ	0.649	**0.843**	0.733
	BiTERM	0.755	0.543	0.632

Таблица 2: Precision, recall и F1-мера для маркировки предложений аспектами ресторанной тематики

Аспект	**Метод**	**Precision**	**Recall**	**F1**
Камера	CATT	**0.860**	0.838	**0.849**
	ATT	0.776	0.829	0.802
	FREQ	0.765	0.744	0.754
	BiTERM	0.604	**0.884**	0.718
Аккумулятор	CATT	**0.867**	0.681	0.763
	ATT	0.840	**0.829**	**0.835**
	FREQ	0.790	0.708	0.747
	BiTERM	0.714	0.742	0.728
Динамики	CATT	**0.786**	0.433	**0.558**
	ATT	0.457	**0.682**	0.547
	FREQ	0.394	0.434	0.413
	BiTERM	0.417	0.613	0.496
Цена	CATT	0.877	0.756	0.813
	ATT	**0.882**	**0.914**	**0.898**
	FREQ	0.874	0.851	0.863
	BiTERM	0.767	0.729	0.748
Память	CATT	**0.722**	0.712	**0.717**
	ATT	0.568	0.933	0.706
	FREQ	0.508	**0.905**	0.651
	BiTERM	0.479	0.847	0.612
Дисплей	CATT	**0.882**	0.557	0.683
	ATT	0.716	**0.747**	**0.731**
	FREQ	0.715	0.536	0.613
	BiTERM	0.668	0.581	0.621
Производительность	CATT	0.518	0.487	0.502
	ATT	0.524	0.522	0.522
	FREQ	**0.546**	0.534	**0.541**
	BiTERM	0.329	**0.750**	0.457

Таблица 3: Precision, recall и F1-мера для маркировки предложений аспектами мобильных телефонов

6 Заключение

В данной магистерской диссертации получены следующие результаты:

1. Проведен обзор современных методов решения задачи извлечения аспектов.
2. По результатам обзора выбраны и программно реализованы методы извлечения аспектов из мнений на русском языке.
3. Подготовлены наборы мнений о двух различных сущностях: ресторанах и мобильных телефонах.
4. Проведено экспериментальное исследование качества работы реализованных методов и проанализированы результаты их работы.

Экспериментальное исследование показало состоятельность предложенных методов для решения задачи извлечения аспектов из текстов мнений о любой предметной области. Код созданного программного модуля размещен на GitHub по ссылке `https://github.com/anick2/master_thesis`

Список литературы

[1] Исследование: влияние отзывов на мнение потребителя [Электронный ресурс]. `https://vc.ru/marketing/91417`.

[2] *Искакова, М. Е.* Подход к автоматическому анализу отзывов о товарах и услугах интернет-магазина / М. Е. Искакова // *Молодой ученый*. — 2023. — № 10 (457). — С. 11–14.

[3] *Большакова, Е. И.* Автоматическая обработка текстов на естественном языке и анализ данных / Е. И. Большакова, К. В. Воронцов, Ефремова Н. Э., Клышинский Э. С., Лукашевич Н. В., Сапин А. С. — Издательство НИУ ВШЭ, 2017. — Pp. 154–168.

[4] *M., Pontiki.* Semeval-2014 task 4: Aspect based sentiment analysis / Pontiki M., H. Papageorgiou, D. Galanis, I. Androutsopoulos, J. Pavlopoulos, S. Manandhar // In Proceedings of the 8th International Workshop on Semantic Evaluation. — 2014. — Pp. 27–35.

[5] *Лукашевич, Н. В.* SentiRuEval: тестирование систем анализа тональности текстов на русском языке по отношению к заданному объекту / Н. В. Лукашевич // *Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной международной конференции «Диалог»*. — 2015.

[6] *Алиева, А. В.* Исследование подходов к аспектному анализу тональности текстов и существующих программных решений / А. В. Алиева. — 2021.

[7] *Turney, P. D.* Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews / P. D. Turney. — 2002. DOI: 10.3115/1073083.1073153.

[8] *Vargas, F.* Aspect Clustering Methods for Sentiment Analysis / F. Vargas // Computational Processing of the Portuguese Language. — Pp. 365–374. DOI: 10.1007/978-3-319-99722-3_37.

[9] *Hu, M.* Mining and summarizing customer reviews / M. Hu, B. Liu // Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — 2004. DOI: 10.1145/1014052.1014073.

[10] *Рой, Д. А.* Методы извлечения аспектных терминов из мнений / Д. А. Рой, Н. Э Ефремова // *Новые информационные технологии в автоматизированных системах.* — 2018.

[11] *Sowjanya, M.* Aspect Based Sentiment Analysis using POS Tagging and TFIDF / M. Sowjanya, K. Srividya // *International Journal of Engineering and Advanced Technology.* — 2019. — no. 8(6). DOI: 10.35940/ijeat.F7935.088619.

[12] *Lun-Wei, K.* Opinion extraction, summarization and tracking in news and blog corpora / K. Lun-Wei, Y. Liang, H. Chen // *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs.* — 2006.

[13] *Hu, M.* Mining Opinion Features in Customer Reviews / M. Hu, B. Liu // AAAI Conference on Artificial Intelligence. — 2004.

[14] *Проноза, Е. В.* Аспектный анализ отзывов о ресторанах для рекомендательных систем е-туризма / Е. В. Проноза, Е. В. Ягунова // Труды XVIII объединенной конференции «Интернет и современное общество». — 2015. — Pp. 130–141.

[15] *Niklas, J.* Extracting opinion targets in a single and cross-domain setting with conditional random fieldss / J. Niklas, I. Gurevych // Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010). — 2010. — Pp. 1035–1045.

[16] *Choi, Y.* Hierarchical sequential learning for extracting opinions and their attributes / Y. Choi, C. Cardie // In Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010). — 2010. — Pp. 269–274.

[17] *Ruder, S.* Deep Learning for Multilingual Aspect-based Sentiment Analysis / S. Ruder, P. Ghaffari, J. G. Breslin // Proceedings of SemEval-2016. — 2016.

[18] *Tamchyna, A.* Recurrent Neural Networks for Sentence Classification / A. Tamchyna, K. Veselovska // Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). — 2016. DOI: 10.18653/v1/S16-1059.

[19] *Khalil, T.* Deep Convolutional Neural Networks for Aspect Category and Sentiment Extraction / T. Khalil, S. R. El-Beltagy // *Cognitive Computation.* — 2016. DOI: 10.1007/s12559-023-10127-6.

[20] *He, R.* An Interactive Multi-Task Learning Network for End-to-End Aspect-Based Sentiment Analysis / R. He, W. Lee, H. Ng // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. — 2019. DOI: 10.18653/v1/P19-1048.

[21] *He, R.* An Unsupervised Neural Attention Model for Aspect Extraction / R. He, W. Lee, H. Ng, D. Dahlmeier // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). — 2017. DOI: 10.18653/v1/P17-1036.

[22] *Tulkens, S.* Embarrassingly Simple Unsupervised Aspect Extraction / S. Tulkens, A. Cranenburgh // Annual Meeting of the Association for Computational Linguistics. — 2020.

[23] *Titov, I.* A Joint Model of Text and Aspect Ratings for Sentiment Summarization / I. Titov, R. McDonald // Annual Meeting of the Association for Computational Linguistics. — 2008.

[24] *Yan, X.* A biterm topic model for short texts / X. Yan, J. Guo, Lan Y. // *Proceedings of the 22nd international conference on World Wide Web.* — 2013. DOI: 10.1145/2488388.2488514.

[25] *Rybakov, V.* Aspect-Based Sentiment Analysis of Russian Hotel Reviews / V. Rybakov, A. Malafeev // CEUR WORKSHOP PROCEEDINGS. — 2018. — Pp. 75–84. DOI: 10.1016/j.procs.2017.09.115.

[26] *Biancofiore, G.* Aspect Based Sentiment Analysis in Music: a case study with Spotify / G. Biancofiore, T. Noia, E. Sciascio, F. Narducci, P. Pastore // SAC '22:

The 37th ACM/SIGAPP Symposium on Applied Computing. — 2022. DOI: 10.1145/3477314.3507092.

[27] *S., Xiong.* Exploiting Capacity-Constrained K-Means Clustering for Aspect-Phrase Grouping / Xiong S., Ji D. // International Conference on Knowledge Science, Engineering and Management. — 2015. DOI: 10.1007/978-3-319-25159-2_34.

[28] NLTK [Электронный ресурс]. `https://www.nltk.org/`.

[29] Морфологический анализатор pymorphy2 [Электронный ресурс]. `https://pymorphy2.readthedocs.org/en/latest/`.

[30] *Mikolov, T.* DistributedRepresentationsofWords and Phrases and their Compositionality / T. Mikolov, Sutskever I., K. Chen, G. Corrado, J. Dean. — 2013.

[31] gensim [Электронный ресурс]. `https://radimrehurek.com/gensim/index.html`.

[32] scikit-learn [Электронный ресурс]. `http://scikit-learn.org/stable/`.

[33] PyTorch [Электронный ресурс]. `https://pytorch.org/`.

[34] *Alsabti, K.* A biterm topic model for short texts / K. Alsabti, S. Ranka, V. Singh. — 1997.

[35] *Mukherjee, A.* Aspect extraction through semi-supervised modeling / A. Mukherjee, B. Liu // Conference: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1. — 2012.

Приложение А. Пример экспертной разметки

На Рисунке 10 представлен пример экспертной разметки, выполненной одним экспертом для ресторанной тематики. Для каждого аспекта эксперт должен быть предложить его название и провести разметку аспектных терминов. Зеленым цветом отмечены термины, подходящие под указанный аспект, красным – неподходящие.

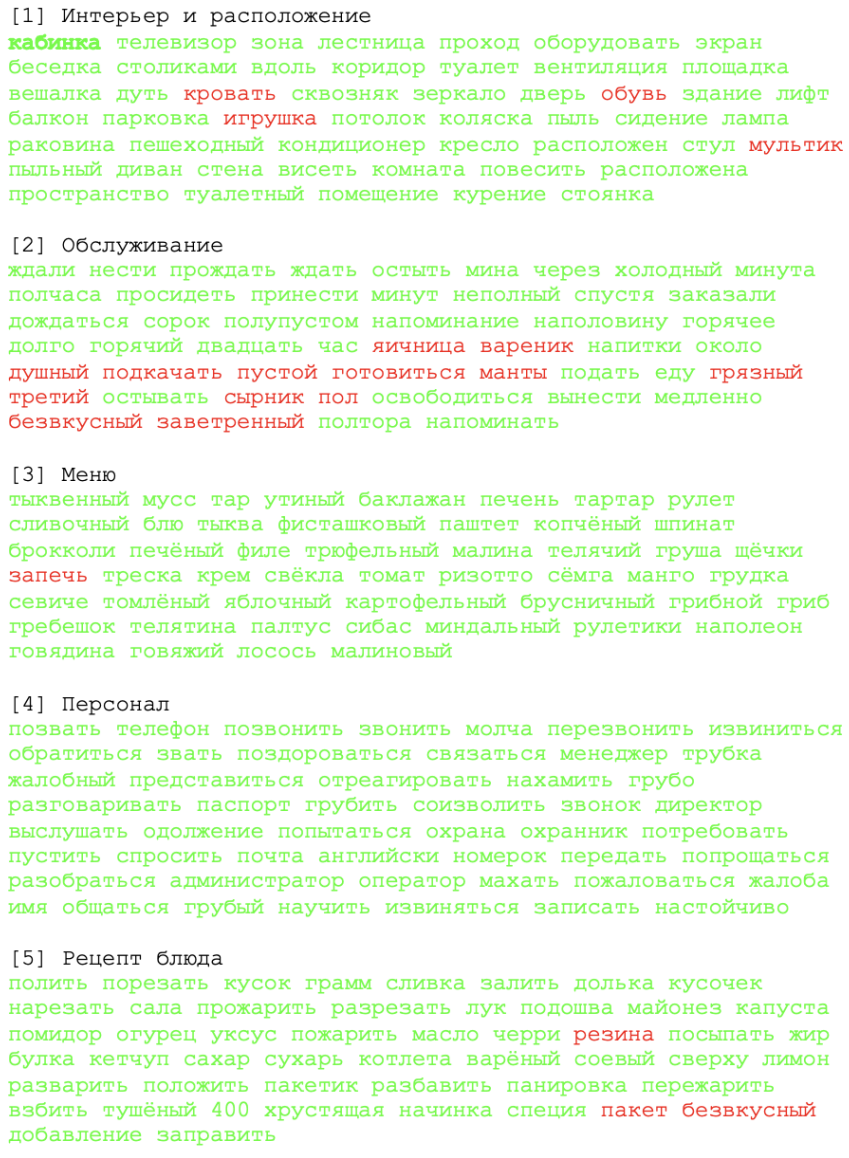

Рис. 10: Пример экспертной разметки для аспектов ресторанной тематики

Приложение Б. Интерпретация аспектов экспертами

На Рисунке 11 представлен результат интерпретации аспектов четырьмя экспертами для ресторанной тематики, на Рисунке 12 – для телефонов. Зеленая заливка ячейки таблицы означает, что более половины экспертов смогли его интерпретировать, следовательно, аспект извлечен верно, красная – менее половины экспертов смогли интеретировать аспект, желтая — ровно половина экспертов смогли интерпретировать аспект.

	FREQ	ATT	BITERM
1	Расположение – Локация –	Интерьер – Пространство Интерьер	Интерьер Обстановка Пространство Интерьер
2	Профессионализм Конфликт Сотрудник Стафф	Обслуживание Ожидание Обслуживание Обслуживание	– – – –
3	– – – –	Меню Меню Еда Меню	– Впечатление – Впечатление
4	– Обстановка Пространство –	Празднование Фуршет Праздник Праздничное мероприятие	Обслуживание Ожидание Обслуживание Обслуживание
5	Еда Еда Блюдо Меню	Сотрудники Персонал Администрация Резерв	– – – –
6	Персонал Персонал Сотрудники Персонал	Общие впечатления Стиль Атмосфера Атмосфера	– – – –
7	Оценка Оценка качества Оценка Оценка	Паттерн посещения Любимое место Локация –	Стоимость Меню Цена Стоимость
8	Напитки Напитки Напиток Напитки	Рецепт Рецепт Готовка Рецепт	Еда Еда Еда Кухня
9	Еда Горячее Основное блюдо Меню	Стилистика Интерьер Стиль Стиль	Общие впечатления Оценка Качества Впечатления
10	Состав Ингредиенты Ингредиенты Ингредиенты	– – Событие Семейное посещение	– – – –

Рис. 11: Интерпретация экспертами аспектов для ресторанов

	K-means	ABAE	BTM
1	– Покупка Покупка Покупка	– Замена – Замена	– – – –
2	Камера Камера Камера Камера	Форм-фактор Дизайн Стиль Внешний вид	Батарея Батарея Зарядка Батарея
3	Настройка Настройки – Перепрошивка	Использование Функционал Мобильный интернет Приложения	– – Сервис –
4	Повреждения Прочность Падение Краш-тест	Состояние Качество покрытия Состояние Корпус	– – – –
5	Операционные системы Модель – Модели	Повреждения Повреждение Поломка Краш-тест	– Покупка Стоимость Стоимость
6	Модели Модели Модели Бренды	Камера Камера Камера Камера	– – – –
7	– – – –	– Глюки Звонок Звонок	Покупка – Покупка Покупка
8	Использование Социальные сети Социальные сети Приложения	Модели Модель Модель Бренды	Повреждения – Поломка Повреждения
9	Производительность Производительность Глюк Глюки	– Приложение – Прошивки	– – Использование Приложения
10	– Магазин Продажа Магазин	– – – –	Модели Модель Модель Бренды

Рис. 12: Интерпретация экспертами аспектов для мобильных телефонов