Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра интеллектуальных информационных технологий

Димов Илья Николаевич

Методы автоматического определения структуры и полемической позиции аргументации

Научный руководитель:
к.ф.-м.н. Добров Борис Викторович

Москва, 2021

Аннотация

Методы автоматического определения
структуры и полемической позиции аргументации

Димов Илья Николаевич

Настоящая работа посвящена исследованию методов решения задачи автоматического извлечения аргументации в текстах.

В работе проводится обзор существующих моделей аргументации и соответствующих подходов ее выделения. Предлагаются методики по адаптации моделей для задач определения структуры и полемической позиции аргументации.

Здесь краткое изложение или повтор заключения.

Содержание

1 Введение

Аргументационная теория, или аргументация, является междисциплинарным исследованием о том, как выводы могут быть достигнуты через череду логических рассуждений. Извлечение аргументации – область науки, стоящая на стыке обработки естественного языка, информационного поиска и непосредственно аргументационной теории.

В общем виде аргумент состоит из утверждения и набора предпосылок, связанных с этим утверждением. Утверждение и предпосылки называются компонентами аргументации. Связи между ними могут выражать не только структуру аргумента, но и тип отношения внутри нее – поддержку или опровержение (атаку). Данные отношения называются полемической позицией аргументации.

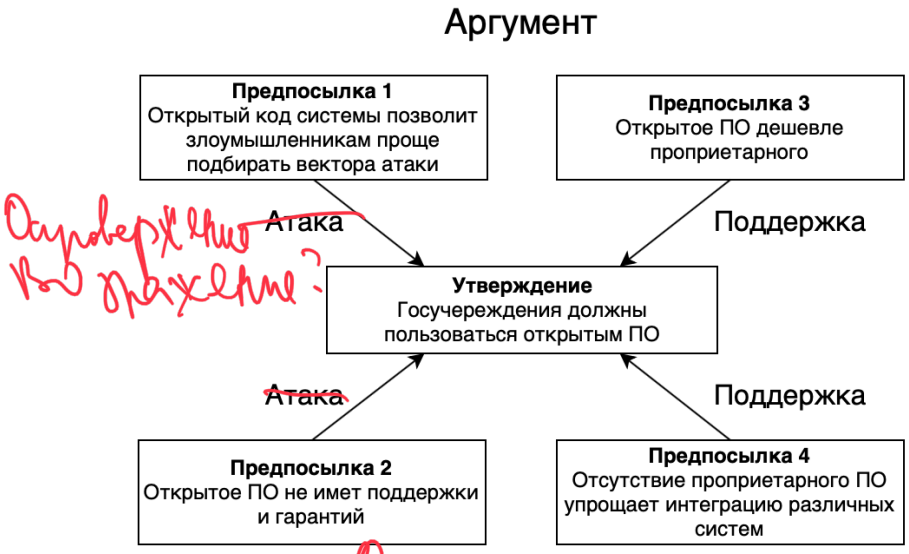

Рисунок 1. Пример аргумента, относящегося к применению открытого ПО.

Помимо вышеуказанных структурных особенностей, связи компонент аргументации могут иметь дополнительно описываться следующими характеристиками:

- Сила – мера того, насколько предпосылки соглашаются с утверждением, темой или опровергают их.
- Общность – мера того, насколько утверждение или предпосылка раскрывает тему, а не ее конкретные примеры, иначе оно превращается в факт.

- Авторитетность - характеристика того, насколько данное утверждение или предпосылка обоснована. Например высказывание ученого о глобальном потепленее авторитетнее высказывания политика.

Аргументация присутствует почти во всех источниках информации: она встречается в бытовых спорах, дебатах, научном дискурсе. Особый вид политической аргументации – пропаганда является ключевым компонентом некоторых интернет-ресурсов.

За последнее десятилетие сильно вырос интерес к автоматическому выделению аргументации [1]. Этому поспособствовал прорыв в области обработки естественного языка, заключающийся создании предобученных нейросетевых моделей с сильными обобщающими способностями [2, 3, 4]. Несмотря на то, что данные модели содержат в себе знание общеизвестных и часто встречаемых фактов, выученных из больших коллекций [5], они не способны применять узкоспециализированую информацию. Для решения данного недостатка используют графы знаний и другие онтологии [6, 7], однако их создание требует трудоемкий и долгий процесс ручной разметки данных экспертами. Автоматическое извлечение аргументации позволяет динамически извлекать специализированную информацию и структурировать ее.

В данной задаче отстутствует устоявшаяся постановка. В научной литературе существует несколько вариантов определения аргументации и соответстующих подходв к ее выделению. Из-за этого отсутствует общепринятый набор тестов и метрик, позволяющих оценить качество системы извлечения аргументации. Следствием отсутствия универсальной постановки задачи является небольшое число исследовательских групп, занимающихся аргументацией, наличие различных по постановке и небольших по размерам обучающих коллекций и почти полное отсутствие развития задачи на языках, отличных от английского.

Проблему плохого покрытия языка можно решать созданием новых или параллельных [8] корпусов, а также методикой межязыкового переноса знаний [9, 10]. Данный подход заключается в обучении мультиязычной модели на одном языке для адаптации полученных знаний на других языкак.

В данной работе описывается исследование методов извлечения структуры и полемической позиции аргументации, а также способы адаптации данных методов для применения на русскоязычных корпусах с помощью подхода межязыкового переноса.

1.1 Актуальность задачи

Извлечение аргументации можно применить во многих прикладных сценариях:

- Аргументация может быть задействована в экспертных системах для получения доводов "за"и "против"относительно какого-либо утверждения.
- Аргументы можно использовать в голосовых помошниках и других системах, использующих базы знаний.
- Выделение аргументации применительно к научным текстам позволит лучше выстроить структуру взаимосвязи между цитируемыми работами.
- Аргументация также может использоваться для оценки убедительности и логичности текстов.
- Аргументация может использоваться в нефактоидных вопросно-ответных системах для ответов на вопросы, требующие объяснений: "Почему было принято следующе решение?"

Один из самых показательных примеров применения аргументации является проект IBM Debater, который способен поддерживать сложный диалог с оппонентом для обоснования определенной точки зрения. В 2019 году система IBM Debater участвовала в споре с финалистом мирового чемпионата по дебатам Харишем Натараджаном, в котором на протяжении 15 минут смогла отстаивать свою точку зрения. Другим примером могут послужить голосовые помошники, получившие широкое распространение в последние годы. В лидирующих системах, таких как Siri, Алиса, Маруся, Alexa и Google Assistant для ответов на вопросы ищутся релевантные документы, которые потом цитируются в качестве ответа, однако они не способны агрегировать информацию из нескольких источников. Добавление аргументации поможет голосовым помошникам систематизировать информацию и обосновывать свои ответы. В качестве примера можно привести работу [11], где по извлеченным примерам аргументации генерируется текст.

Из всего вышесказанного можно сделать вывод, что выделение аргументации с помощью методов глубокого обучения является перспективной и актуальной задачей. Научная ценность данной работы заключается в исследовании и систематизации текущих

подходов к задаче извлечения аргументации, применении новых моделей и предоставлении замеров результатов их работы. Отдельной проблемой является и работа с множеством небольших, разнящихся в постановке задачи корпусов. Следствием является и проблема построения системы извлечения аргументации для русского языка.

В работе приводится обзор и анализ существующих работ по извлечению аргументации, предлагаются новые подходы, основанные на переносе знаний (transfer learning), а также исследуется возможность межязыкового переноса знаний для получения модели, работающей на русском языке. Дополнительно проводится интерпретация работы модели: исследуется зависимость полученных результатов в зависимости от структурных особенностей компонент аргументации, таких как наличие отрицаний, сильно окрашенных слов или антонимии.

2 Постановка задачи

Целью данной работы является исследование существующих методов извлечения структуры и полемической позиции аргументации и разработка новых модельных подходов.

Перед дальнейшими определениями необходимо пояснить термен "компоненты аргументации". Определение и набор компонент разнятся от работы к работе, но в общем случае можно выделить следующие сущности:

- Фокус - центральный объект обсуждения.
- Тема - противоречивое утверждение относительно фокуса.
- Утверждение - фраза или предложение, поддерживающие или опровергающее тему.
- Предпосылка - фраза или предложение, поддерживающие или опровергающие или тему или предложение, основанное на фактах, а не на убеждении.

Под структурой аргумента подразумевается набор компонент аргументации и отношения связности или релевантности между ними. Например для темы "запрет ядерной энергии"утверждение "ядерная энергия вредна для окружающей среды"является релевантным, а утверждение "человеческий глаз воспринимает около миллиона оттенков"релевантным не является. Отношение релевантности является бинарным отношением между компонентами аргументации.

Задача определения полемической позиции заключается в определении типа связи между релевантными компонентами аргументации. Релевантное теме утверждение может как соглашаться с темой, так и ее опровергать. Задача классификации связей между релевантными компонентами аргументации в два класса поддержки и атаки называется определением полемической позиции. Как видно из определения, полемическая позиция применима исключительно к релевантным компонентам аргументации и представляет собой два взаимоисключающих класса связей.

Для достижения поставленной в работе цели необходимо решить следующие подзадачи:

1. Произвести обзор существующих решений и корпусов, выделить наиболее подходящие постановки.
2. Воспроизвести избранные работы или получить новые базовые решения.
3. Предложить или адаптировать новые модельные подходы для извлечения аргументации.
4. Провести анализ полученной модели с целью интерпретации полученных результатов.

3 Обзор подходов к извлечению аргументации

3.1 Обзор корпусов

Задача извлечения аргументации решается в различных постановках. Общей идей является выделение компонент аргументации и определение связей между ними. Наиболее серьезные работы с применением машинного обучения, посвященные теме извлечения аргументации, датируются 2014 годом. В работах [12, 13] предлагается корпус аргументации на 2500 примеров. В качестве компонент аргументации рассматриваются три сущности:

- Тема - короткая противоречивая фраза, определяющая центральный объект обсуждения (фокус).
- Контекстно-зависимое утверждение - краткая фраза, напрямую подтверждающая или опровергающая тему.
- Контекстно-зависимая предпосылка - участок текста, напрямую поддерживающий контекстно-зависимое утверждение в рамках данной темы.

Задача поставлена следующим образом - по заданной теме и коллекции текстов из Википедии необходимо выделить контекстно-зависимые утверждения и предпосылки. Этап выделения компонент аргументации разделен на две подзадачи. Сначала предлагается выделить предложения, содержащие компоненты аргументации, после чего выделить непрервыне участки текста (фразы), которые являются непосредственно предпосылкой или утверждением. Таким образом решается задача выделения структуры аргументации. Как следует из определений, в данном корпусе отсутствует связь противоречия (атаки, противопоставления) между компонентами, поэтому отсутствует возможность выучить полемическую позицию аргумента. Стоит отметить, что задача выделения непрервыного участка текста в качестве компонент аргументации требует сложной разметки, из-за чего в следующих работах отказываются от этого этапа и решают задачу извлечения компонент на уровне предложений.

Дополнительно поставлена задача классификации предпосылок в три класса - прецедентные (Anecdotal), т.е. основанные на каком-либо случае; эксперные (Expert),

т.е. высказывание авторитетного источника; исследовательские (Study), т.е. результаты какого-либо научного исследования.

В работе [14] предлагается новый корпус IBM Evidence Search размером 5700 примеров извлечения контекстно-зависимых предпосылок по темам. В данном корпусе происходит изменение постановки задачи. В предыдущих работах предпосылки имели связь с утверждениями, которые в свою очередь имели связь с темами, теперь предпосылки связаны напрямую с темами. В данном корпусе решается задача определения релевантности предпосылки по отношению к теме, т.е. решается задача извлечения структуры аргументации. В разметке отсутствует полемическая позиция предпосылки - поддержка или опровержение темы.

Предлагается два корпуса - SLD (strong labeled data, т.е. данные полученные с помощью человеческой экспертизы), а так же WLD (weak labeled data, т.е. данные, полученные автоматическими правилами), однако корпус WLD не предоставлен. В работе исследуется зависимость качества модели определения релевантности между предпосылкой и темой в зависимости от соотношения корпусов WLD и SLD. Показано, что использование данных, полученных автоматически повышает итоговое качество модели на несколько пунктов с 70% до 72-74% точности классификации, усредненной по темам.

В работе [15] предлагается корпус UKP Sentential Argument Mining Corpus. Рассматриваются пары сущностей "тема"и "утверждение определяется наличие связи и ее тип. В отличие от предыдущих работ, тема не является противоречывым высказыванием, а вместо этого представляет собой центральный объект обсуждения (фокус). Утверждение является предложением, которое может как поддерживать или опровергать фокус, так и не иметь к нему отношения, что позволяет решать и задачу определения структуры аргументации и полемической позиции. Примеры: тема - "ядерная энергия поддерживающее утверждение - "ядерная энергия уменьшает парниковый эффект". В корпусе представлено 25000 пар тем и утверждений по 8 темам.

В работе исследуется способность моделей работать на новых темах, т.е. обобщать свои знания на домены, которые раньше не были выучены моделью. Несмотря на ограниченный набор тем, данный корпус одним из первых позволяет решать одновременно и задачу извлечения структуры аргументации и определения полемической позиции. Корпус отсутствует в открытом доступе.

В работе [16] подробнее описываются техники автоматического получения корпусов

на примере задачи определения релевантности утверждения по отношению к теме. В качестве основного способа предлагается использовать правила, основанные на определенных лексических маркерах. Например предполагается, что часто встречаются шаблоны вида <**someone**>**argued that**<**claim**>. В корпусе отсутствует полемическая позиция и решается только задача определения структуры аргументации. В качестве замеров эффективности обученная на подобном корпусе модель замеряется на UKP Sentential Argument Mining Corpus и получает результаты сравнимые с моделями, предложенными в оригинальной работе.

Исследование [17] полностью посвещено задаче определения полемической позиции утверждений по отношению к теме. В корпусе IBM Claim Stance предоставлено 2400 пар тем и утверждений связанных отношением поддержки или атаки. Отсутствуют примеры утверждений не связанных с темами, поэтому отсутствует возможность извлечь структуру аргументации.

Отдельно стоит отметить и задачу определения качества аргументации, которой посвещен целый ряд работ [18, 19, 20]. В данной задаче необходимо определить качество утверждения по отношению к теме. Решение данной задачи позволяет отсортировать утверждения, чтобы использовать наиболее подходящие в целевой задаче. Сущестует несколько постановок данной задачи: присваивание утверждению некоторой величины от 0 до 1, отражающей его качество, или попарное сравнение двух утверждений с выбором наилучшего.

В данных работах рассматриваются поддерживающие утверждения релевантные теме, поэтому невозможно обучить модель на этих корпусах ни определению полемической позиции, ни структуре аргументации, однако нельзя не отметить полезность подобного ранжирования в полномасштабной системе извлечения аргументации.

В работе [21] предлагается два мультиязычных корпуса для извлечения аргументации. В первом корпусе ArgsEN было предоставлено 30000 пар тем и утверждений, у которых размечена полемическая позиция и качество аргумента. Помимо этого предоставлен корпус EviEN, состоящий из 35000 пар предпосылок, в котором размечена и полемическая позиция и релевантность.

Свой подход к аргументации развивается и в исследовательской команде Webis. В корпусе Webis Debate 16 [22] предлагается классифицировать предложения в два класса - нейтральные и содержащие утверждение. В данной текстовой коллекции утверждения

не привязаны к теме или любой другой компоненте аргументации; данная постановка не позволяет выделить ни структуру аргументации, ни полемическую позицию внутри нее.

Нельзя не внести в обзор и такую задачу как логический вывод (Textual Entailment, Natural Languge Inference). Данная задача напрямую не занимается извлечением аргументации, но задачи имеют схожую постановку: требуется по тексту и гипотезе понять отношение между ними. Видов отношений 3: нейтральное отношение, поддержка, противоречие. Примеры:

```
Текст: A man inspects the uniform of a figure in some East Asian country.
Гипотеза: The man is sleeping
Отношение: Contradiction

Текст: An older and younger man smilin
Гипотеза: Two men are smiling and laughing at the cats playing on the floor
Отношение: Neutral

Текст: A black race car starts up in front of a crowd of people
Гипотеза: A man is driving down a lonely road.
Отношение: Contradiction
```

Задача NLI интересна тем, что как и в задачах извлечения аргументации присутствует несколько компонент (текст и гипотеза), между которыми необходимо найти связь. Вместо оригинальной постановки, в которй требовалось классифицировать отношение между двумя предложениями в 3 класса, можно факторизовать задачу в две подзадачи: определить релевантна ли гипотеза тексту, после чего определить при условии релевантности поддерживает ли или опровергается текст гипотезой. Данная формулировка очень похожа на этапы извлечения аргументации, а именно на извлечение структуры и последующее определение полемической позиции.

Существует несколько корпусов для решения подобной задачи: Stanford NLI [23], Multi-Genre NLI, TERRa. В отличие от корпусов аргументации, размер корпусов логического вывода измеряется не в тысячах примеров, а в сотнях тысяч. Более того, корпуса для этой задачи встречаются на нескольких языках, в том числе и русском.

Как было сказано в определении предпосылки, важно, чтобы предпосылка не являлась убеждением или верованием, а имела под собой веские основания. Во всех описанных выше корпусах, где предлагалось работать с предпосылками, предполагалось, что эта закономерность заложена в данные. В соревновании **CheckThat!** [24] при конференции CLEF с 2020 задача верификации предпосылок решается целенаправленно. В данном соревновании представлены следующие дорожки, собранные из данных социальной сети Twitter:

- **Значимость (Check Worthiness)** - даны сообщения пользователей, необходимо отранжировать их по значимости, т.е. таким образом, чтобы на первых местах оказались те сообщения, которые необходимо верифицировать в первую очередь.
- **Выделение предпосылок (Claim Retrieval)** - дано утверждение, представляющее из себя некое сообщение из твиттера, и набор уже верифицированных утверждений. Необходимо отранжировать верифицированные утверждения, чтобы понять, можно ли с их помощью подтвердить целевое утверждение. Данный трек доступен только на арабском языке.

Похожим образом построен и корпус FEVER (Fact Extraction and VERification) [25]. Необходимо по набору утверждений и статей из википедии найти участки текста, подтверждающие целевое утверждение.

В работе [26] предоставляется Internet Argument Corpus v2. В данной коллекции размечены пары сообщений из интернет-форумов. Сообщения разбиты на пары-вопрос-ответ и размечены по следующим критериям:

1. Согласие или несогласие - число от -5 до 5, отражающее полемическую позицию ответа к вопросу.
2. Эмоциональность или факты - число от -5 до 5, отражающее насколько в ответе используются факты, а не эмоции и убеждения отвечающего.
3. Сарказм - число от 0 до 1, отражающее число разметчиков, посчитавших ответ саркастичным, а не серьезным.

Как видно из описания данный корпус предоставляет определенную разметку вебдискурса, из которой можно определенными эвристиками получить корпус для извлечения структуры и полемической позиции аргументации.

В более ранних работах встречались небольшие корпуса из очень узких тематических областей. В работах [27, 28] предоставлена схожая разметка в юридических документах и научных статьях соответственно. Разметка представляет с собой связи между предложениями внутри текстов - как предложения поддерживают друг друга, чтобы итоговый текст был связным и убедительным.

Дополнительно стоит отметить и примеры политической аргументации, содержащиеся в корпусе трека SemEval 2020 Task 11 Propaganda detection in news articles [29]. В данном соревновании предлагается выделить участки новостных статей, содержащих аргументацию, а также классифицировать их в такие приемы, как ложная аналогия, подмена тезиса, эксплуатация двусмысленных выражений, сведение аргументации к универсально осуждаемой теме и другим приемам убеждения оппонента. Пропаганду также можно свести и к примерам "ложной"аргументации, так как она скорее нацелена убеждение читающего, а не на установление истины.

3.2 Обзор существующих решений

Было создано несколько систем для извлечения аргументации, работающих с различными постановками задачи.

Первая рассматриваемая система MARGOT [30] придерживается определений контекстно-зависимых утверждений и предпосылок из работы [12]. По заданной теме к тексту последовательно применяется несколько моделей - модель определения предложений, содержащих компоненты аргументации, и модель выделения компоненты внутри предложения, после чего компоненты аргументации связываются отношением поддержки. Предложенные модели основаны на методах машинного обучения над признаками, выделенными из синтаксиса предложений.

Система MARGOT решает задачу извлечения аргументации в одной из самых ранних постановок, от которой отказались в последующих работах. В данной постановке отсутствует отношение противоречия, что не позволяет полностью определить ни структуру, ни полемическую позицию аргументации. Более того в данной постановке дополнительно происходит выделение компонент аргументации внутри предложений, как коротких непрерывных участков фраз. От данной подзадачи впоследствии отказались в пользу решения задачи на уровне предложений.

Targer [31] - система направленная на анализ текстов с целью изучения их аргу-

ментационной структуры. В системе предлагается использовать подход классификации последовательностей слов (Sequence Tagging), подобный задаче распознавания именованных сущнностей, для разбиения текста на фразы и выделения связей между ними. Данная система работает в рамках определенного текста и не способна связывать произвольные компоненты аргументации, выделенные из различных источников.

ArgumentText [32] - данная система отличается от предыдущих постановкой задачи. Вместо поиска участков текста, которые являются компонентами аргументации и выделения отношений между ними, производится поиск предложений, напрямую поддерживающих или опровергающих тему. ArgumenText представляет собой многоступенчатую систему извлечения аргументации, в рамках которой решается и задача кластеризации утверждений по теме, поиск и классификация релевантных документов.

3.3 Выводы к разделу 3

Как видно из обзора работ существует множество подходов к задаче извлечения аргументации, как и существует несколько постановок. Также есть и ряд подзадач, таких как верификация и ранжирование утверждений по качеству или определение противоричивости темы.

Основными критериями для выбора корпусов служили следующие критерии:

- Возможность напрямую решать или задачу определения структуры или определения полемической позиции аргументации. В то время как сущетсвует множество различных постановок задачи извлечения аргументации, а так же вспомогательных задач, в рамках данной работы было решено сфокусироваться на основных задачах извлечения аргументации.

- Возможность сравнить свое модельное решение с заявленными метриками качества. Некоторые работы или не имеют заявленных чисел или имеют невоспроизводимые результаты (замеры качества с помощью экспертов), или используют в обучении дополнительные закрытые наборы данных, которые не позволяют адекватно сравнить полученные системы.

По этим причинам были выбраны следующие корпуса:

1. Новостной корпус пропаганды из соревнования SemEval [29].

2. Корпус для построения структуры аргументации IBM Evidence Search из работы [14]

3. Корпус для определения полемической позиции ArgsEN и корпус для одновременного определения полемической позиции и структуры аргументации EviEN из работы [21].

4. Корпус из задачи логического вывода SNLI [23] для исследования применимости закономерностей, выученных на задаче логического вывода в задаче извлечения аргументации.

5. Корпус задачи логического вывода TERRA на русском языке

⇒ Обзор корпусов большой,
а обзор решений – маленький

4 Методы решения задачи

В рамках проблемы извлечения аргументации необходимо решить две подзадачи: выделение структуры аргумента и определение полемической позиции. Обе эти задачи являются задачами классификации. Формально даны два текста (предложения) $A = (a_1, a_2, ..., a_n)$ и $B = (b_1, b_2, ..., b_m)$, где a_i, b_j - слова. Необходимо для пары компонент аргументации A, B определить класс отношения $Y = (y_1, ..., y_k)$, т.е. смоделировать функцию (P обозначает вероятность):

$$f(A, B, y_i) = P(y_i|A, B)$$

В задаче определения структуры аргументации присутствует два класса: отношение релевантности и нерелевантности (нейтральное). В задаче определения полемической позиции также два одношения: поддержка и опровержение (~~атака~~).

В задаче определения пропаганды присутствует 14 классов пропаганды, постановка задачи остается похожей: дан новостной текст $X = (x1, ..., x_n)$, дан участок $x_i, ..., x_i + k$, необходимо классифицировать данный сегмент в один из классов $Y = (y_1, ..., y_14)$:

- Сильно окрашенная лексика. Пример: ".. a lone lawmaker's childish shouting."
- Навешивание ярлыков и имен. Пример: "Republican congressweasels"
- Повторение. Заключается в повторении тезиса для закреплении концепции в голове у слушателей.
- Преувеличение или приуменьшение. Пример: "Democrats bolted as soon as Trump's speech ended in an apparent effort to signal they can't even stomach being in the same room as the president"
- Сомнения. Пример: "Is he ready to be the Mayor?"
- Взывание к страхам и предрассудкам. Пример: "stop those refugees; they are terrorists"
- Сбор под флагом, т.е. взывание к аудитории на почве общей какой-либо общей идентичности (национальности, пола и т.п.). Пример: "entering this war will make us have a better future in our country"

- Упрощение причины, т.е. упор на какую-либо первопричину в проблеме с множеством факторов. Пример: "If France had not declared war on Germany, World War II would have never happened."

- Слоганы - любые яркие фразы, взывающие к эмоциям. Пример: "Make America great again!"

- Диктатура - представление ограниченного выбора в задаче, где существует большее число вариантов решение проблемы. Пример: "There is no alternative to war."

- Терминирующие размышления клише - фразы дискредитирующие критичекское размышление по теме обсуждения. Пример: "никто не идеален"или "нельзя изменить человеческую натуру"

- Whataboutism - перевод темы для занятия более высокой моральной позиции. Пример: "А у вас права ущемляют!".

- Сведение вопроса к заведомо осуждаемой теме. Пример: "Only one kind of person can think this way: a communist!"

- Красная сельдь - представление нерелевантной информации для отвращения внимания.

- Стадность - захват мнения через представление позиции, как поддерживаемой большинством. Пример: "Would you vote for Clinton as president? 57% say yes."

- Нарочитая расплывчатость и неясность.

- Соломенный человек - подстановка похожего утверждения вместо утверждения оппонента с его последующим опровержением.

В задаче извлечения пропаганды дополнительно стоит задача нахождения участков текста, являющихся пропагандой, т.е. для текста $X = (x_1, ..., x_n)$ определить бинарный набор меток $Y = (y_1, ..., y_n)$, где $y_i = 1$, если слово x_i входит в сегмент с пропагандой, иначе $y_i = 0$. Для решения проблемы необходимо построить модель:

$$f(X, Y) = P((y_1, y_2, ..., y_n)|(x_1, x_2, ..., x_n))$$

4.1 Современные подходы к обработке естественного языка

За последнее десятилетие область обработки текстов значительным образом преобразилась. Тексты хуже поддаются обработке различными нейросетевыми алгоритмами из-за их дискретной структуры. Ранние работы, основаные на классических метотдах машинного обучения используют методики Bag of words или Tf-Idf для векторизации текста. Данные подходы не способны улавливать сложные закономерности, так как не учитывают порядок и внутреннюю структуру текста, поэтому их зачастую дополняют информацией о синтаксической структуре предложения и другуми экспертными признаками.

В 2013 году с появлением word2vec [2] началось стремительное развитие применения нейросетей в обработке текстов. Основная идея данного подхода заключается в выучивании векторов слов для моделирования дистрибутивной гипотезы: слова, встречающиеся в схожих контекстах, имеют близкие значения.

Долгое время стандартом моделей была одна из векторизаций (word2vec, glove, fasttext) дополненная рекуррентной нейросетью (LSTM, GRU) [33] и финальным слоем специфичным для задачи. В данном подходе рекуррентные нейросети выучивыают контекстные закономерности в рамках определенной задачи.

Со временем начали появляться модели общего назначения для задачи восприятия естественного языка (Natural Language Understanding). Данные модели не требуется обучать с нуля, вместо этого они проходят длительное предобучения на общей задаче моделирования языка [34]: для последовательности $X = (x_1, x_2, .., x_k)$ слов из словаря $V = (v_1, v_2, ..., v_n)$ моделируется следующая закономерность:

$$f(X, v_i) = P(x_{k+1} = v_i | (x_1, ..., x_k))$$

В настоящий момент большие предобученные модели основанные на архитектуре трансформер являются доминирующими почти во всех областях обработки текстов. Особенностью данной архитектуры является полный отказ от рекуррентности в пользу механизма внимания.

Механизм внимания пришел из задачи машинного перевода, где главным подходом была архитектура Seq2Seq [35]. Данная архитектура состоит из энкодера, сжимающего предложение в вектор, и декодера, разжимающего данный вектор в текст на дру-

гом языке. Одна из главных заявленнных проблем - данный подход плохо работает на длинных предложениях. Это связано с тем, что при сжатии в вектор фиксированного размера теряется очень большой объем информации.

Прорыв произошел с разработкой механизма внимания [36]. Данный механизм позволяет сопоставлять вектор предложения в декодере с соответствующими состояними в энкодере. Каждому состоянию энкодера, соответствуещему определенному токену, на каждом шаге при декодировке ставится в соответствие вес, после чего состояния энкодера усредняются с данными весами и используются для расширения внутреннего состояния декодера.

Архитектура трансформер [37] основывается на механизме внимания. В рамках нейросетевой модели механизм внимания применяется от каждого слова предложения ко всем словам предложения. С последовательным применением нелинейности данный механизм применяется несколько раз на каждом слое нейросети, после чего на выходе получается векторное представление для каждого слова в предложении, насыщенное контекстуальной информацией.

Одной из самых популярных предобученных архитектур является BERT [3]. Данная модель обучалась на коллекции Book Corpus задаче маскированного языкового моделирования и предсказания следующего предложения. Задача маскированного языкового моделирования заключается в предсказании одного и более слов в предложении. Таким образом после обучения модель имеет внутри себя знания о структуре языка, о мире, поэтому ее легко адаптировать для различных вариаций задачи понимания естественного языка.

Это подтверждается успехом подобных моделей почти во всех современных задачах естественного языка и особом соревновании GLUE Benchmark. Данное соревнование содержит в себе ряд задач на понимание языка, которые при небольшом дообучении решаются моделями подобными BERT на уровне человека или выше.

В рамках данной работы было выбрано использовать нейросетевые подходы на основе предобученных моделей-трансформеров, так как данные модели показывают наилучшие результаты на большинтстве задач обработки естественного языка.

4.2 Модели для извлечения аргументации

Для извлечения аргументации необходимо решить подзадачи построения структуры и определения полемической позиции аргументации. В рамках рассмотренных корпусов компонентами аргументации являются темы и утверждения или предпосылки. Стоит дополнительно отметить, что не у всех представленных текстовых коллекций присутствуют какие-либо базовые решения и замеры качеств.

В работе, сопровождающей корпус IBM Claim Stance [17] предложенная модель представляет собой метод опорных векторов (Support Vector Machine) над векторизацией мешка слов (Bag of words).

В работе [21] для определения полемической позиции, структуры и качества аргумента предлагается система, основанная на предобученной модели $BERT$, однако сама архитектура не раскрывается.

Отсутствие предложенных базовых решений приводит к необходимости использования заимствованных или альтернативных подходов. Для классификации отношений между двумя парами компонент аргументации предлагается модель $BERT_{basic}$. В рамках данной модели входы, представляющие из себя две компоненты аргументации объединяются в один текст, разделенные специальным символом **[SEP]**. После этого входная последовательность проходит сквозь слои модели, в результате чего механизм внимания, интерпретируя закономерности между двумя текстами, выдает сжатый вектор. Далее данный вектор подается в линейный слой, который решает задачу классификации.

Данная модель вдохновлена оригинальной методикой предобучения модели $BERT$, в которой аналогичным образом решалась задача предсказания следующего предложения (Next sentence prediction) - необходимо было определить, является ли одно предложение продолжением другого.

Стоит отметить, что сложность работы модели квадратична относительно длины входной последовательности, и, возможно, данный подход не был бы эффективен в сценарии, где необходимо было бы быстро извлечь аргументацию из большого количества текстов. Альтернативой первой модели служит модель $BERT_{vectorized}$, в которой предлагается независимо пропустить две компоненты аргументации через слои предобученной модели, после чего сконкатенировать их векторные представления и передать для классификации в линейный слой. За счет уменьшения длины входных данных умень-

шается и время работы модели, однако компоненты аргументации векторизуются без информации друг о друге, что может привести к потере качества.

В работе [38] предлагается подход улучшения качества модели за счет добавления интерпретируемости. Данная модель $BERT_{SE}$ представляет собой дополненую модель $BERT_{basic}$: в конце после прохождения всех слоев трансформера для каждого слова на выходе имеется вектор, отображающий контекстуальную информацию предложения внутри текста, т.е. каждая входная последовательность $X = (x_1, ..., x_n)$ отображается в последовательность векторов $H = (h_1, ..., h_n), h_i = (h_{i_1}, ..., h_{i_k},$ где k - размерность модели. В обычной задаче классификации берется выход на первом слове h_1, т.е. на спецсимволе начала предложения **[CLS]**. В модели $BERT_{SE}$ предлагается взять всевозможные участки текста вида $span(i, j), i \leq j, |span(i, j)| = k$ и с помощью линейного слоя проставить каждому участку текста некое число s_q, отражающее важность данного участка текста в финальной классификации. После этого s_q превращается в множители для взвешенной суммы участков текста, финальное векторное представление t для классификации считается следующим образом:

$$\alpha_i = \frac{\exp(s_i)}{\sum_{j=1}^{k} \exp(s_j)}$$

$$t = \sum_{q=1, i \leq j}^{k} \alpha_q * span(i, j)$$

Далее взвешенное представление всех участков t подается в линейный слой для решения задачи классификации. Множители α_i могут использоваться для интерпретации модели, так как они отображают самый важный для классификации непрерывный участок текста.

Дополнительн для исследования применимости знаний, полученных на корпусе SNLI предлагается обучить модель $BERT_{basic}$ на задаче логического вывода и исследовать следующие сценарии: применимость модели без дообучения на корпусах извлечения аргументации (Zero Shot) и исследование предобучения модели на корпусе SNLI для дообучения целевой задаче извлечения аргументации.

4.3 Модели для обнаружения пропаганды

Задача извлечения пропаганды является новой задачей, к которой не было применено никаких существующих подходов. В дорожке по детекции, где было необходимо выделить участки текста, являющиеся пропагадной, было предложено начать с базовых моделей из работы [39]. Первая модель $LSTM$ основывается на одноименной архитектуре рекуррентных нейросетей. На вход модели подается текст, представляемый последовательностью слов $X = (x_1, .., x_n)$, после чего данный текст проходит через LSTM-слои и преобразуется в контекстные вектора $H = (h_1, ..., h_n), h_i = (h_{i_1}, ..., h_{i_k}$. Каждому слову на основе полученных векторов независимо ставится метка 0 (отсутствие пропаганды) или 1 (пропаганда) с помощью применения линейного слоя. Данные метки составляют список меток $Y = (y_1, ..., y_n)$, который однозначно определяет все участки пропаганды в тексте.

Модификацией данной модели является $LSTM_{CRF}$. Предлагается устранить недостатки предудыщей модели за счет замены финального слоя на слой Conditional Random Field (CRF). Данный слой выучивает возможные переходы и моделирует всю последовательнсть Y одновременно.

В качестве более сложных моделей предлагается использовать предобученную модель $BERT$ для векторизации входной последовательности и аналогично с предыдущими моделями в качестве выходных слоев испытать и линейный слой и CRF.

Также было решено позаимствовать архитектуру $LaserTagger$ из работы [40], так как данная модель зарекомендовала себя для классификации длинных последовательностей. Основная идея модели заключается в том, чтобы аналогично задаче машинного перевода, используя архитектуру "encoder-decoder"поставить каждому слову соответствующую метку.

Для улучшения работы модели $LaserTagger$ было предложено два дополнительных механизма Label Smoothing и Teacher Forcing. Техника label smoothing заключается в преобразовании меток класса из 0 и в более "мягкие такие как 0.2 и 0.8. Данная методика позволяет сгладить резкие изменения градиентов на ошибках модели. Teacher forcing заключается в подаче авторегрессионной модели не только правильрных меток, но и предсказанных ею неверных меток. Таким образом модель становится более устойчивой к ошибкам и исследуюет во время обучение больше различных сценариев.

Во второй дорожке требовалось для каждого сегмента пропаганды определить, к

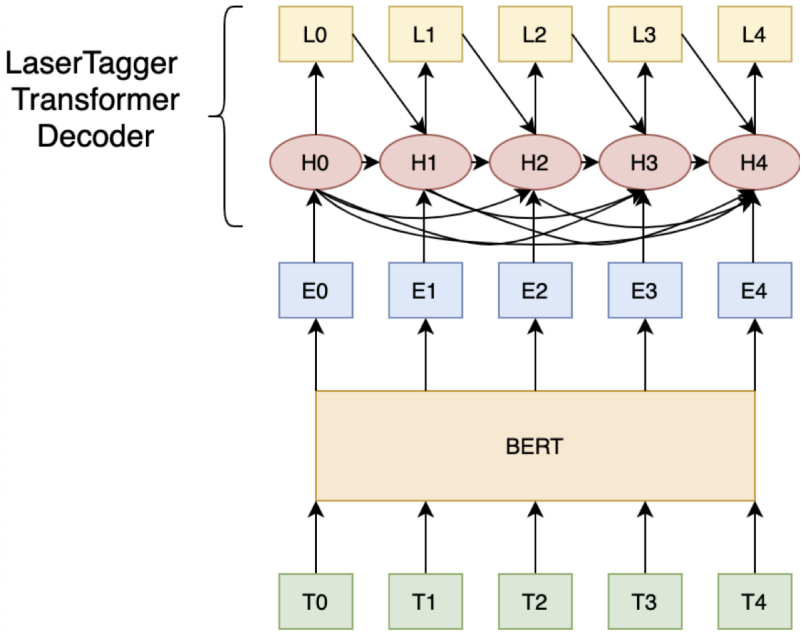

Рис. 2: Архитектура LaserTagger.

каким классам он относится. Задача поставлена на уровне предложений, однако решать ее как классификацию предложений невозможно: в одном предложении может содержаться более одного участка пропаганды. Для этого было решено адаптировать модельный подход $RBERT$ [41], заключающийся в обрамлении нужных участков текста спецсимволами. Таким образом модели выделяется нужная фраза, что позволяет при необходимости работать с разными участками текста в рамках одного предложения.

4.4 Методы для межязыкового переноса знаний

В области обработки естественного языка наиболее распространены англоязычные корпуса. Английский язык получает наибольшее покрытие задачами, однако проверка работоспособности моделей на других языках из-за этого остается плохо исследованной проблемой.

Методов получить модель для другого языка несколько, направления межязыкового переноса активно развивается в настоящее время. Самым простоым способом является создание параллельного корпуса. Использование профессиональных переводчиков затратно и треюует много времени, поэтому достаточно часто для первичной проверки гипотез прибегают к автоматизированным средствам перевода.

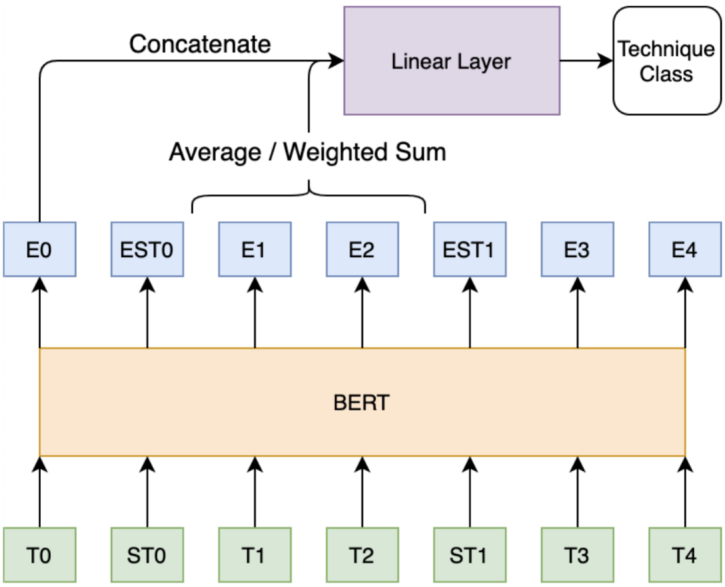

Рис. 3: архитектура R-BERT.

Второй подход заключается в адаптации мультиязычной модели, т.е. модели предобученной на корпусе из различных языков. Считается, что для представлений одних и тех же слов на разных языках модель имеет схожие векторные представления, которые "замораживаются т.е. не обучаются дальше. После этого данная модель дообучается на конкретной задаче на тех языках, на которых данная задача доступна, а качество замеряется уже на целевом языке.

4.5 Выводы

В результате обзора современных нейросетевых подходов были выбраные следующие модели для задачи извлечения аргументации:

- Модель $BERT_{basic}$, т.к. она является базовой вариацией модели классификации, основанной на больших предобученных моделях.
- Модель $BERT_{vectorized}$ для исследования важности применения механизма внимания между компонентами аргументации.
- Модель $BERT_{SE}$ для исследования интерпретируемости результатов.

Для извлечения пропаганды были предложены две модели:

- Модель, основанная на рекуррентной нейросети $LSTM$, так как это хорошее базовое решение, зарекомендовавшее себя в задачах классификации последовательностей.

- Модель $LaserTagger$, так как она зарекомендовала себя на задачах классификации длинных последовательностей.

- Модель $RBERT$ для классификации участков текста, так как она позволяет выделить интересующий участок пропаганды.

Для исследования переноса знаний на русский язык было выбрано перевести корпус IBM Claim Stance с помощью автоматизированного сервиса Google Translate и применить подход с заморозкой входных слоев мультиязычной модели $BERT_{multilingual}$ для дальнейшего замера качества на переведенном корпусе.

5 Програмная реализация

В качестве языка для реализации работы[1] был выбран Python 3.6. Данный язык позволяет быстро разрабатывать простые сервисы и имеет ряд библиотек для работы с нейросетевыми моделями. В качестве фреймворка для глубокого обучения используется Pytorch 1.6.0 и библиотека AlenNLP 1.1.0.

Кодовую базу можно разделить на несколько модулей:

- Модуль обработки корпусов (800 строк) - данный код отвечает за обработку текстовых коллекций в форматы для обучения. Исходные форматы корпусов в форматах CSV, JSON, XML, MySQL dump были преобразованы в человекочитаемый формат JSONL.

- Модуль моделей (500 строк). В данном модуле написаны нейросетевые архитекутры и классы для взаимодействия с ними.

- Модуль обучения (500 строк). В данном модуле написаны циклы обучения, подсчет метрик, интеграции с системой отслеживания экспериментов[2].

- Веб-сервис (250 строк) - реализован стенд для удобного взаимодействия с моделями на микрофреймворке Flask 1.1.1.

Выбор библиотек Pytorch и AllenNLP обоснован их простотой и возможностью быстрого прототипирования нейросетевых архитектур. Выбор микрофреймворка Flask для веб-сервиса был сделан из-за возможности легко создать веб-приложение на несколько страниц для взаимодействия с моделями.

Обучение происходило удаленно на сервере DataCrunch[3] с видеокартами Tesla V100 с 16Гб видеопамяти. Процесс обучения в зависимости от корпуса занимал от 30 минут до 8 часов.

Для исследования эффективности работы моделей извлечения аргументации было создано веб-приложение. Веб-приложение содержит в себе одну страницу для взаимодействия с моделями. На странице предоставлены поля для компонент аргументации

[1] https://github.com/hawkeoni/Thesis
[2] https://wandb.ai/hawkeoni
[3] https://datacrunch.io/

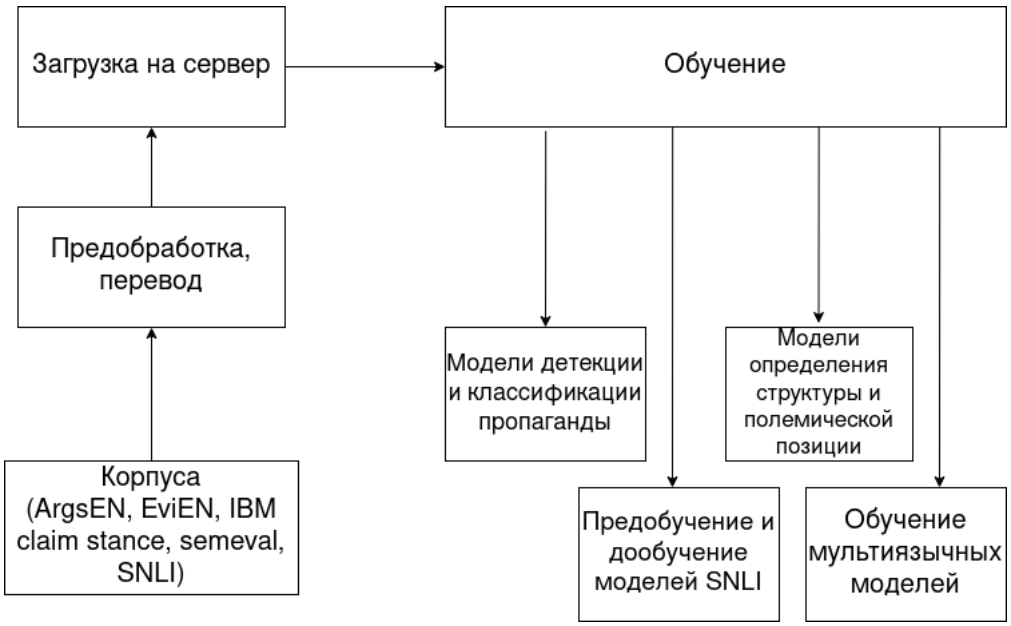

Рис. 4: Процесс обучения модели.

и выбора моделей. После заполнения полей и выбора моделей на странице с помощью языка JavaScript отображается результаты работы моделей определения структуры аргументации (релевантности) и полемической позиции (поддержка или атака).

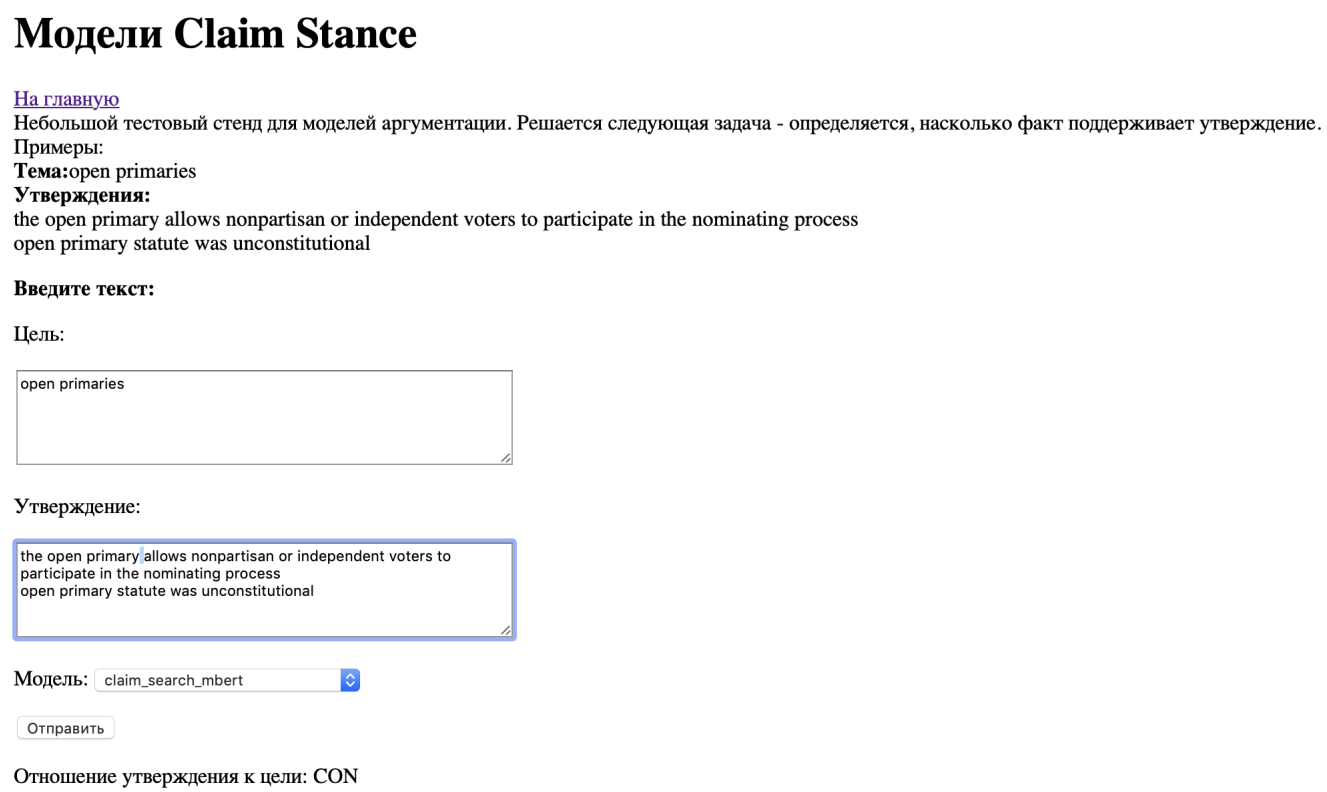

Рисунок 5. Интерфейс сервиса.

6 Экспериментальное исследование

6.1 Эксперименты на корпусе IBM Evidence Search

Для первых экспериментов для извлечения структуры аргументации был выбран корпус IBM Evidence Search из работы [14]. В рамках данной текстовой коллекции необходимо найти связь между темой и предпосылкой. Примеры:

```
Тема: We should fight illegal immigration
Предпосылка: A paper in the peer reviewed Tax Lawyer journal from the American
Bar Association asserts that illegal immigrants contribute more in taxes
than they cost in social services
Тип отношения: релевантно

Тема: We should abandon coal mining
Предпосылка: In particular, Daintree was the first Government geologist for
North Queensland discovering gold fields and coal seams for future exploitation.
Тип отношения: нерелевантно
```

В данном корпусе представлено 5700 пар тем и предпосылок по 118 темам, 35 из которых отведены в качестве тестовой выборки. Примеров положительного класса, т.е. пар релевантных темы и посылки и в тестовой и в обучающей части около 40%. Дополнительно тестовая часть этого корпуса была переведена на русский язык для замеров эффективности межязыкового переноса.

Также рассматривается русскоязычный корпуса TERRa. В данном корпусе стоит задача логического вывода: необходимо по паре текста и гипотезы понять, можно ли вывести гипотезу из текста. Данная постановка похожа на постановку определения релевантности, поэтому было решено исследовать влияние обучения на корпусе TERRa на решение задачи определения структуры аргументации. Пример данных из корпуса TERRa:

```
Текст: Женщину доставили в больницу, за ее жизнь сейчас борются врачи.
Гипотеза: Женщину спасают врачи.
Тип отношения: гипотезу можно вывести из текста
```

```
Текст: Представитель внешнеполитического ведомства отметил, что злоумышленника
задержали местные власти.
Гипотеза: Злоумышленник напуган.
Тип отношения: гипотеза не следует из текста.
```

Модель	Макро-Точность	Микро-Точность
$BERT_{en}$	**82.4**	**81.1**
$BERT_{mult}$	81.6	80.5
$IBMBiLSTM$	-	72
$IBMBiLSTM_{ext}$	-	76
$BERT_{mult}$	77.5	76.4
$BERT_{terraZS}$	66.7	60.3
$BERT_{comb}$	75.9	73.9

Таблица 1: Результаты на тестовом корпусе IBM Evidence Search. Горизонтальной линией отделены результаты на англоязычной и русскоязычной версиях.

В качестве метрики в исходной работе используется микро-точность по темам. В качестве моделей авторы [14] предлагали двунаправленную LSTM над векторизацией слов GloVE. Данная модель обозначена как $IBMBiLSTM$. Авторы также исследовали влияние внешних данных, собранных автоматически на качество работы модели. Модель, использующая внешние данные называется $IBMBiLSTM_{ext}$. Автоматически собранный корпус не выложен в открытый доступ, поэтому исследовать его влияние на предложенные в данной дипломной работе модели не представляется возможным.

В качестве предложенных моделей для обучения на английском корпусе было решено взять архитектуру $BERT_{base}$, т.е. модель, одновременно обрабатывающую обе компоненты аргументации. В таблице 1 моделью $BERT_{en}$ называется модель на архитектуре $BERT_{base}$, дообученная с весов англоязычной версии $BERT$. Модели $BERT_{mult}$, $BERT_{terraZS}$, $BERT_{comb}$ дообучены с весов мультиязычной версии оригинальной модели $BERT$.

Как видно из таблицы использование англоязычной модели $BERT_{en}$ дает заметно лучшее качество, чем модели, основанные на архитектуре $BiLSTM$, даже с использова-

нием внешних данных. На один пункт точности уменьшается качество мультиязычной модели $BERT_{mult}$ по сравнению с англоязычной, но качество все еще заметно выское.

Во второй половине таблицы отображены результаты работы модели $BERT_{mult}$ на русскоязычной версии корпуса. Происходит заметное падение качества на 5 пунктов, однако модель все еще извлекает аргументацию на уровне, сопоставимом с наилучшим решением.

Модель $BERT_{terraZS}$ была обучена на корпусе TERRa и применена в сценарии Zero Shot (без дообучения) на корпусе IBM Evidence Search. Модель показывает ненулевое качество. Дальнейшей попыткой применить корпус логического вывода TERRa является модель $BERT_{comb}$. Было решено объединить корпуса TERRa и IBM Evidence Search, чтобы понять как влияет корпус TERRa на задачу извлечения структуры аргументации. Полученная модель оказывается хуже аналогичной модели $BERT_{mult}$, обученной исключительно на англоязычных данных.

6.2 Эксперименты на корпусах ArgsEN и EviEN

В новой работе [21] предоставлено 2 корпуса, посвященных аргументации: ArgsEN и EviEN. В корпусе ArgsEN на 30000 примеров размечена полемическая позиция между темами и утверждениями:

```
Тема: We should abandon marriage
Утверждение: \"marriage\" isn't keeping up with the times.  abandon the old
thinking and bring something that incorporates all unions - not just those with a
man and woman.
Полемическая позиция: поддержка

Тема: We should prohibit flag burning
Утверждение: a flag is only really a peace of cloth and doesn't actually hurt
anybody.
Полемическая позиция: противоречие (атака)
```

В корпусе EviEN похожим образом размечены отношения между темами и предпосылками (35000 пар), дополнительно размечена и релевантность. Таким образом корпус

EviEN является единственным корпусом, на котором можно решать одновременно задачи извлечения струкутуры и определения полемической позиции. Примеры:

```
Тема: We should ban the sale of violent video games to minors
Предпосылка: Justice Thomas, in his dissent, considered that historically,
the Founding Fathers "believed parents to have complete authority over
their minor children and expected parents to direct the
development of those children
Структура: релевантно
Полемическая позиция: противоречие

Тема: We should ban the sale of violent video games to minors
Предпосылка: while owning guns is a legal right in most countries,
the illegal trade in guns continues to fuel conflict
Структура: нерелевантно
Полемическая позиция: нейтральная
```

Также дополнительно рассматриваится и корпус для логического вывода SNLI. Данный корпус состоит из 570000 пар текстов и гипотез, необходимо определить поддерживается, опровергается гипотеза текстом или из текста нельзя сделать выводы о гипотезе. Примеры:

```
Тема: A person on a horse jumps over a broken down airplane.
Гипотеза: A person is outdoors, on a horse.
Отношение: Подтверждение
```

На корпусе ArgsEN проводились замеры качества определения полемической позиции аргументации, отображенные в таблице 2.

Авторы оригинальной работы представляют результаты моделей $IBMBERT_{en}$ и $IBMBERT_{17L}$. Архитектуры моделей не описываются, известно, что первая модель училась на английском корпусе, а вторая модель училась на неопубликованном корпусе из 17 языков. Вторая модель показывает наилучшие результаты и включена в сравнение для предоставление как лучшая обученная модель, однако предоставленные в данной дипломной работе модели справедливо сравнивать лишь с первой моделью, т.к. они учились исключительно на английском языке.

Модель	Макро-F1
$IBMBERT_{en}$	89.3
$IBMBERT_{17L}$	*91.5*
$BERT_{basic}$	90.1
$BERT_{vectorized}$	75.2
$BERT_{SE}$	**90.3**
$BERT_{snliZS}$	55.0
$BERT_{snliFT}$	87.9

Таблица 2: Результаты на тестовом корпусе ArgsEN в задаче определения полемической позиции утверждения относительно темы. Метрика - макроусредненная F1 по темам.

Модели $BERT_{basic}$, $BERT_{vectorized}$, $BERT_{SE}$ начинали обучение с англоязычных весов модели $BERT$ и представляют собой архитектуры, описанные в 4 главе. Модель $BERT_{vectorized}$ из-за отсутствия перекрестного механизма внимания, вызванного независимой обработкой компонент аргументации показывает низкий результат. Значительно лучше ведет себя модель $BERT_{basic}$, способная смоделировать сложные завивимости между темой и утверждением, так как они подаются в модель одновременно. Модель со слоем интерпретации $BERT_{SE}$ показывает результат немного превосходящий базовую модель.

Дополнительно использовалась модель обученная на корпусе SNLI. В оригинальном корпусе SNLI присутствует 3 отношения: поддержка, атака и нейтральное отношение. Т.к. в данном корпусе нет нерелевантных пар темы и утверждения, из предсказаний модели выбирается наиболее вероятный класс из поддержки или атаки. В сценарии без дообучения (Zero Shot) модель $BERT_{snliZS}$ показывает низкий результат в 55 пунктов точности на задаче бинарной классификации. Дообученная модель $BERT_{snliFT}$ показывает также показывает более низкий результат, чем модель $BERT_{basic}$. Данный результат совпадает со схожим экспериментом на корпусах IBM Evidence Search и TERRa.

На корпусе EviEN тоже присутствует возможность выделять полемическую позицию, однако авторы работы [21] не предоставляют для сравнения никаких модельных результатов. Модели, аналогичные моделям для корпуса ArgsEN представлены в таблице 3.

Модель	Макро-F1
$BERT_{basic}$	64.8
$BERT_{vectorized}$	51.8
$BERT_{SE}$	67.6

Таблица 3: Результаты на тестовом корпусе EviEN в задаче определения структуры и полемической позиции предпосылки относительно темы. Метрика - макроусредненная F1 по темам.

Как отмечалось ранее, на корпусе EviEN поставлена вторая задача определения релевантности предпосылки по отношению к теме. Авторы корпуса не предоставляют своих модельных метрик. Метрики моделей из главы 4 предоставлены в таблице 4:

Модель	Макро-F1
$BERT_{basic}$	68.5
$BERT_{vectorized}$	64.3
$BERT_{SE}$	72.8

Таблица 4: Результаты на тестовом корпусе EviEN в задаче определения структуры аргументации (релевантности). Метрика - макроусредненная F1 по темам.

Результаты на корпусе ArgsEN превосходят модели, предложенные в работе [21] по метрикам. На всех корпусах лучше всего себя показывает модель $BERT_{SE}$ со слоем интерпретации, а хуже всего модель $BERT_{vectorized}$.

6.3 Обнаружение пропаганды

Корпус пропаганды состоял из ручной разметки новостных статей за 2019 год. Всего было обработано 438 статей, содержащих 21230 предложений, из которых 7485 содержали пропаганду.

В качестве первого трека были даны новостные статьи и было необходимо выделить участки текста, содержащие пропаганду.

В качестве базовых моделей были предложены архитектуры на основе рекуррентной нейросети BiLSTM. Модель $BiLSTM_{glove+charlstm}$ использует для векторизации эм-

Модель	F1	Precision	Recall
$BiLSTM_{glove+charlstm}$	34.9	30.4	41.1
$BiLSTM_{ELMO}$	34.5	32.7	36.7
$BERT_{linear}$	40.8	35.7	47.7
$BERT_{crf}$	36.2	30.1	45.4
$BERT_{lstm+crf}$	41.4	36.3	48.2
$LaserTagger$	42.0	38.4	46.4
$LaserTaggertf$	45.1	42.3	48.2
$LaserTaggertf + ls$	46.1	40.6	53.3
$LasterTagger_{TF+LS}$	44.6	55.6	37.3
$Hitachi$	51.5	56.5	47.3

Таблица 5: Результаты на задаче выделения пропаганды. В качестве метрики используется посимвольная F1-мера. Чертой отделены замеры на валидационной выборке и тестовой.

беддинги GloVe и буквенный рекуррентный слой BiLSTM. Данный подход позволяет обрабатывать слова, которые не встречались в исходном словаре GloVe. Второй вариацией данной модели является векторизация на основе буквенной нейросети ElMO [34]. Из таблицы 5 видно, что два этих подхожа не имеют большой качественной разницы.

Далее ведется работа с англоязычной версией нейросетевой модели $BERT$. Модель, в которой слова векторизуются $BERT$ и подаются в линейный слой называется $BERT_{linear}$, что дает большой прирост качества относительно базовых моделей. Минусом данной модели является независимое предсказание последовательности меток. Для решения данной проблемы использовалась нейросеть $BERT_{crf}$, в которой слой Conditional Random Field одновременно оценивает всю последовательность меток, однако данный метод не дает улчшения, а наоборот ухудшает результат.

В качестве продолжения идеи научить модель связанно предсказывать метки была принята архитекутра $LaserTagger$. Данная модель работает аналогично моделям машинного перевода - исходная последовательность проходит векторизацию моделью $BERT$, после чего каждому слову в соответствие ставится метка. В качестве модификаций данной модели были предложены методики Teacher Forcing и Label Smoothing,

комбинация которых и вошла в финальную модель с качеством 44.6 посимвольной F1-меры.

Модель	Микро-F1
$BERT_{CLS}$	57.3
$R-BERT$	59.2
$R-BERT_{ft}$	58.4
$R-BERT_{w}$	59.0
$R-BERT_{w+ft}$	59.0
$Ensemble$	60.6
$Ensemble$	58.2
$ApplicaAI$	62.07

Таблица 6: Результаты на задаче классификации пропаганды. В качестве метрики используется микро-F1 по классам пропаганды. Чертой отделены замеры на валидационной выборке и тестовой.

Для классификации участков пропаганды была выбрана модель $R-BERT$. Для работы данной модели необходимо обрамить участки с пропагандой спецсимволами, чтобы указать модели, какой участок текста является наиболее важным. В базовом варианте для классификации бралось усреднение контекстных векторных представлений выделенного сегмента. В модели $R-BERT_w$ модель выучивала веса для участков текста и брала взвешенную сумму их векторных представлений для классификации. В модели $R-BERT_{ft}$ исходная модель $BERT$ дообучалась на доменном корпусе новостей, что дало небольшой прирост качества. В финальной модели использовался ансамбль всех моделей, собранный в голосующий классификатор.

Стоит отметить, что в корпусе прослеживается сильное смещение тем между тестовой выборкой и тренировочной, валидационной. Из-за этого метрики всех участников на валидационной выборке на 10-15% выше, чем итоговые результаты на тесте. Команды Hitachi и ApplicaAI применили модели, схожие с $BERT_{linear}$, однако использовали внешние данные и методики доразметки данных моделью для радикального увеличения тренировочной выборки. По результатам[42] участия было занято 7 место из 36 в первом треке и 6 из 31 во втором треке.

7 Заключение

В рамках данной работы были получены следующие результаты:

- Проведен обзор существующих методов и корпусов для извлечения аргументации.
- Обучены системы, основанные на больших предобученных языковых моделях, для извлечения аргументации.
- Реализован базовый вариант переноса знаний на русский язык для задачи извлечения структуры аргументации.
- Исследована возможность использования предобучения на задаче логического вывода для улучшения результатов на задачах извлечения аргументации.
- Реализована система для обнаружения пропаганды в новостных ресурсах.
- Реализовано базовое веб-приложение для взаимодействия с моделями.

В данной работе исследованы задачи автоматического выявления аргументации на примере извлечения структуры и определении полемической позиции аргументации. Несмотря на отсутствие крупных корпусов, на которых исследовательские группы целеноправленно сравнивали бы свои модели, были выделены избранные коллекции. На коллекциях ArgsEN [21] в задаче определения полемической позиции аргументации был улучшен результат англоязычной модели с 89.3 пунктов макро F1-меры до 90.3; на коллекции IBM Evidence Search [14] для определения структуры аргументации также улучшен результат с 76 пунктов микро точности до 81.1. На коллекциях, где отсутствовали авторские замеры были также предоставлены показатели качества: на корпусе EviEN на задаче определения полемической позиции - 67.6, на задаче определения структуры аргумента (релевантности) - 72.8.

На корпусе IBM Evidence Search было проведено исследование возможности использования мультиязычной модели для переноса знаний на русский язык. Для этого тестовая выборка была переведена на русский язык. Модель показывает падение точности с 80.5 пунктов для англоязычного корпуса до 76.4 пунктов, что отражает ожидаемое ухудшение качество при смене языка. Тем не менее качество этой модели достигает современного уровня извлечения аргументации.

Дополнительно на корпусе IBM Evidence Search была исследована возможность применения закономерностей, выученных моделью на корпусе логического вывода. Использование корпуса TERRa в сценарии без дообучения показывает возможность применения подобного решения на задаче определения структуры аргументации. Обучение одновременно логическому выводу и структуре аргументации приводит к ухудшению качества, полученного исключительно на корпусе аргументации, что говорит о непригодности использования корпуса TERRa в подобном сценарии. Аналогичные испытания проводились и для корпуса ArgsEN на задаче определения полемической позиции и корпусе SNLI для логического вывода. В этом эксперименте дообученная модель также привела к ухудшению качества по сравнению с моделью, использующей исключительно корпус аргументации.

Дополнительно исследована задача извлечения пропаганды - частного вида аргументации. Предоставлены модели, занявшие 6 и 7 место в научном семинаре Semeval 11 [42].

Для взаимодействия с моделями предоставлен веб-сервис, позволяющий для пар компонент аргументации оценить релевантность и полемическую позицию.

В качестве будущего развития работы можно предложить:

- Дополнение моделей внешними знаниями для моделирования более сложных зависимостей.
- Использование полученных аргументов для создания карт аргументации или генерации текста.
- Создание русскоязычной коллекции для решения задачи извлечения аргументации.

Список литературы

[1] *Lippi, Marco.* Argumentation mining: State of the art and emerging trends / Marco Lippi, Paolo Torroni // *ACM Transactions on Internet Technology (TOIT).* — 2016. — Vol. 16, no. 2. — Pp. 1–25.

[2] Efficient estimation of word representations in vector space / Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean // *arXiv preprint arXiv:1301.3781.* — 2013.

[3] Bert: Pre-training of deep bidirectional transformers for language understanding / Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova // *arXiv preprint arXiv:1810.04805.* — 2018.

[4] Language models are unsupervised multitask learners / Alec Radford, Jeffrey Wu, Rewon Child et al. // *OpenAI blog.* — 2019. — Vol. 1, no. 8. — P. 9.

[5] Language models as knowledge bases? / Fabio Petroni, Tim Rocktäschel, Patrick Lewis et al. // *arXiv preprint arXiv:1909.01066.* — 2019.

[6] *Sorokin, Daniil.* Modeling semantics with gated graph neural networks for knowledge base question answering / Daniil Sorokin, Iryna Gurevych // *arXiv preprint arXiv:1808.04126.* — 2018.

[7] *Chen, Jifan.* Multi-hop question answering via reasoning chains / Jifan Chen, Shih-ting Lin, Greg Durrett // *arXiv preprint arXiv:1910.02610.* — 2019.

[8] *Fishcheva, Irina.* Cross-Lingual Argumentation Mining for Russian Texts / Irina Fishcheva, Evgeny Kotelnikov // International Conference on Analysis of Images, Social Networks and Texts / Springer. — 2019. — Pp. 134–144.

[9] XCOPA: A Multilingual Dataset for Causal Commonsense Reasoning / Edoardo Maria Ponti, Goran Glavaš, Olga Majewska et al. // *arXiv preprint arXiv:2005.00333.* — 2020.

[10] Zero-shot cross-lingual classification using multilingual neural machine translation / Akiko Eriguchi, Melvin Johnson, Orhan Firat et al. // *arXiv preprint arXiv:1809.04686.* — 2018.

[11] *Schiller, Benjamin.* Aspect-controlled neural argument generation / Benjamin Schiller, Johannes Daxenberger, Iryna Gurevych // *arXiv preprint arXiv:2005.00084*. — 2020.

[12] A benchmark dataset for automatic detection of claims and evidence in the context of controversial topics / Ehud Aharoni, Anatoly Polnarov, Tamar Lavee et al. // Proceedings of the first workshop on argumentation mining. — 2014. — Pp. 64–68.

[13] Show me your evidence-an automatic method for context dependent evidence detection / Ruty Rinott, Lena Dankin, Carlos Alzate et al. // Proceedings of the 2015 conference on empirical methods in natural language processing. — 2015. — Pp. 440–450.

[14] Will it blend? blending weak and strong labeled data in a neural network for argumentation mining / Eyal Shnarch, Carlos Alzate, Lena Dankin et al. // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). — 2018. — Pp. 599–605.

[15] *Stab, Christian.* Cross-topic argument mining from heterogeneous sources using attention-based neural networks / Christian Stab, Tristan Miller, Iryna Gurevych // *arXiv preprint arXiv:1802.05758*. — 2018.

[16] Towards an argumentative content search engine using weak supervision / Ran Levy, Ben Bogin, Shai Gretz et al. // Proceedings of the 27th International Conference on Computational Linguistics. — 2018. — Pp. 2066–2081.

[17] Stance classification of context-dependent claims / Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo et al. // Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. — 2017. — Pp. 251–261.

[18] A large-scale dataset for argument quality ranking: Construction and analysis / Shai Gretz, Roni Friedman, Edo Cohen-Karlik et al. // Proceedings of the AAAI Conference on Artificial Intelligence. — Vol. 34. — 2020. — Pp. 7805–7813.

[19] Automatic Argument Quality Assessment–New Datasets and Methods / Assaf Toledo, Shai Gretz, Edo Cohen-Karlik et al. // *arXiv preprint arXiv:1909.01007*. — 2019.

[20] Are you convinced? Choosing the more convincing evidence with a Siamese network / Martin Gleize, Eyal Shnarch, Leshem Choshen et al. // *arXiv preprint arXiv:1907.08971*. — 2019.

[21] Multilingual argument mining: Datasets and analysis / Orith Toledo-Ronen, Matan Orbach, Yonatan Bilu et al. // *arXiv preprint arXiv:2010.06432*. — 2020.

[22] Webis-Debate-16 / Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen et al. — Zenodo, 2016. — . `https://doi.org/10.5281/zenodo.3251804`.

[23] A large annotated corpus for learning natural language inference / Samuel R. Bowman, Gabor Angeli, Christopher Potts, Christopher D. Manning // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP). — Association for Computational Linguistics, 2015.

[24] Overview of CheckThat! 2020: Automatic identification and verification of claims in social media / Alberto Barrón-Cedeno, Tamer Elsayed, Preslav Nakov et al. // International Conference of the Cross-Language Evaluation Forum for European Languages / Springer. — 2020. — Pp. 215–236.

[25] Fever: a large-scale dataset for fact extraction and verification / James Thorne, Andreas Vlachos, Christos Christodoulopoulos, Arpit Mittal // *arXiv preprint arXiv:1803.05355*. — 2018.

[26] Internet argument corpus 2.0: An sql schema for dialogic social media and the corpora to go with it / Rob Abbott, Brian Ecker, Pranav Anand, Marilyn Walker // Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). — 2016. — Pp. 4445–4452.

[27] *Palau, Raquel Mochales*. Argumentation mining: the detection, classification and structure of arguments in text / Raquel Mochales Palau, Marie-Francine Moens // Proceedings of the 12th international conference on artificial intelligence and law. — 2009. — Pp. 98–107.

[28] *Ronzano, Francesco*. Dr. inventor framework: Extracting structured information from scientific publications / Francesco Ronzano, Horacio Saggion // International conference on discovery science / Springer. — 2015. — Pp. 209–220.

[29] SemEval-2020 task 11: Detection of propaganda techniques in news articles / Giovanni Da San Martino, Alberto Barrón-Cedeno, Henning Wachsmuth et al. // Proceedings of the Fourteenth Workshop on Semantic Evaluation. — 2020. — Pp. 1377–1414.

[30] *Lippi, Marco.* MARGOT: A web server for argumentation mining / Marco Lippi, Paolo Torroni // *Expert Systems with Applications.* — 2016. — Vol. 65. — Pp. 292–303.

[31] Targer: Neural argument mining at your fingertips / Artem Chernodub, Oleksiy Oliynyk, Philipp Heidenreich et al. // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. — 2019. — Pp. 195–200.

[32] ArgumenText: Argument classification and clustering in a generalized search scenario / Johannes Daxenberger, Benjamin Schiller, Chris Stahlhut et al. // *Datenbank-Spektrum.* — 2020. — Vol. 20, no. 2. — Pp. 115–121.

[33] *Hochreiter, Sepp.* LSTM can solve hard long time lag problems / Sepp Hochreiter, Jürgen Schmidhuber // *Advances in neural information processing systems.* — 1997. — Pp. 473–479.

[34] Deep contextualized word representations / Matthew E. Peters, Mark Neumann, Mohit Iyyer et al. // Proc. of NAACL. — 2018.

[35] *Sutskever, Ilya.* Sequence to sequence learning with neural networks / Ilya Sutskever, Oriol Vinyals, Quoc V Le // *arXiv preprint arXiv:1409.3215.* — 2014.

[36] *Bahdanau, Dzmitry.* Neural machine translation by jointly learning to align and translate / Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio // *arXiv preprint arXiv:1409.0473.* — 2014.

[37] Attention is all you need / Ashish Vaswani, Noam Shazeer, Niki Parmar et al. // *Advances in neural information processing systems.* — 2017. — Vol. 30. — Pp. 5998–6008.

[38] Self-Explaining Structures Improve NLP Models / Zijun Sun, Chun Fan, Qinghong Han et al. // *arXiv preprint arXiv:2012.01786.* — 2020.

[39] Neural architectures for named entity recognition / Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian et al. // *arXiv preprint arXiv:1603.01360.* — 2016.

[40] Encode, tag, realize: High-precision text editing / Eric Malmi, Sebastian Krause, Sascha Rothe et al. // *arXiv preprint arXiv:1909.01187*. — 2019.

[41] *Wu, Shanchan.* Enriching pre-trained language model with entity information for relation classification / Shanchan Wu, Yifan He // Proceedings of the 28th ACM International Conference on Information and Knowledge Management. — 2019. — Pp. 2361–2364.

[42] *Dimov, Ilya.* NoPropaganda at SemEval-2020 Task 11: A Borrowed Approach to Sequence Tagging and Text Classification / Ilya Dimov, Vladislav Korzun, Ivan Smurov // Proceedings of the Fourteenth Workshop on Semantic Evaluation. — Barcelona (online): International Committee for Computational Linguistics, 2020. — . — Pp. 1488–1494. https://www.aclweb.org/anthology/2020.semeval-1.194.