**Министерство науки и высшего образования Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего образования**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»**

**(Университет ИТМО)**

**Факультет программной инженерии и компьютерной техники**

**КУРСОВАЯ РАБОТА**

**Тема:** Совместный анализ эмоций и тональности на основе текста MELD

**Работу выполнила** Ян Цзяфэн **группы** P4123

(фамилия, имя) (номер группы)

**Руководитель** Махныткина Олеся Владимировна

(фамилия, имя, отчество)

**Работа защищена** " " 2023 г.

**c оценкой**

**Подписи членов комиссии:**

**САНКТ-ПЕТЕРБУРГ, 2023**

СОДЕРЖАНИЕ

[1. Введение 3](#_Toc138206913)

[2. Аналитический обзор 5](#_Toc138206914)

[2.1 Contextual Information and Commonsense Based Prompt for Emotion Recognition in Conversation 5](#_Toc138206915)

[2.2 CoMPM: Context Modeling with Speaker’s Pre-trained Memory Tracking for Emotion Recognition in Conversation 6](#_Toc138206916)

[2.3 DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation 7](#_Toc138206917)

[2.4 Supervised Prototypical Contrastive Learning for Emotion Recognition in Conversation 7](#_Toc138206918)

[2.5 Multimodal Representations Learning Based on Mutual Information Maximization and Minimization and Identity Embedding for Multimodal Sentiment Analysis 8](#_Toc138206919)

[2.6 Knowledge-Interactive Network with Sentiment Polarity Intensity-Aware Multi-Task Learning for Emotion Recognition in Conversations 8](#_Toc138206920)

[2.7 Joint learning on sentiment and emotion classification 9](#_Toc138206921)

[2.8 Joint Sentiment and Emotion Classification with Integer Linear Programming 10](#_Toc138206922)

[2.9 Multimodal Emotion Recognition with Auxiliary Sentiment Information 11](#_Toc138206923)

[2.10 Emotion helps Sentiment: A Multi-task Model for Sentiment and Emotion Analysis 11](#_Toc138206924)

[2.11 Сравнительная таблица 12](#_Toc138206925)

[2.12 Вывод 13](#_Toc138206926)

[3. Эксполаторный анализ корпуса 15](#_Toc138206927)

[4. Baseline 18](#_Toc138206928)

[5. Усложнение моделей 23](#_Toc138206929)

[6. Заключение 31](#_Toc138206930)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 32](#_Toc138206931)

# Введение

В последние годы многие исследователи обратили внимание на распознавание эмоций и тональности. Способность распознавания эмоций и тональности очень полезны для взаимодействия человека с компьютером. Технология распознавания эмоций и тональности используются во многих средствах массовой информации, таких как аудио, видео, мимики, текстовые данные и т.д. Среди них текстовые данные являются очень популярным носителем, а их небольшой объем делает их наиболее подходящими для передачи через Интернет. Кроме того, из-за разнообразия и сложности текстовых данных это позволяет пользователям Интернета обмениваться идеями, мнениями и эмоциями, используя только текст. Поэтому распознавание эмоций и тональности на основе текстовых данных очень ценны.

Стоит отметить, что классификация эмоций и классификация тональности — это две разные задачи. Из-за зависимости между метками эмоций и метками тональности совместное обучение может эффективно собирать общую информацию о эмоциях и тональности, чтобы достичь цели, приносящей пользу обеим задачам одновременно.

Целью этой работы является разработка и обучение модели нейронной сети для совместного анализа эмоций и тональности на основе текста MELD для повышения производительности моделей классификации эмоций.

Задачи, в соответствии с целью работы, приведены ниже:

1. Обзор и анализ существующих подходов к извлечению признаков текста и нейронных сетей, используемых для классификации эмоций и тональности.
2. Эксполаторный анализ корпуса.
3. Разработка и обучение традиционных классификаторов для эмоций и тональности.
4. Реализация нейронных сетей для совместного анализа эмоций и тональности.
5. Проведение экспериментальных исследований и оценка эффективности разработанной модели на основе выбранных критериев.

# Аналитический обзор

Для выполнения данной работы необходимо выполнить аналитический обзор существующих исследований, связанных с этой темой. Проведенный анализ показал, что количество исследований, связанных с анализом эмоций и тональности на основе текста MELD [1], пока относительно невелико. Далее будет описаны 10 источников, посвященных схожей тематике.

## 2.1 Contextual Information and Commonsense Based Prompt for Emotion Recognition in Conversation

В статье Jingjie Yi и т.д. «Contextual Information and Commonsense Based Prompt for Emotion Recognition in Conversation» [2] предлагается модель ERC CISPER, которая объединяет контекстную информацию и здравый смысл, связанные с высказываниями в разговоре, в непрерывную подсказку для повышения производительности ERC. На рисунке 1 представлена архитектура CISPER. Она состоит из трех основных этапов: извлечение признаков, генерация подсказок и предсказание эмоций. Обучаемые вложения псевдотокенов принимаются в качестве непрерывной подсказки для Pre-trained Languge Model (PLM), которая смешивает два вида важной информации. Одним из них является контекстуальная информация в разговоре, а другим — логический вывод, связанный с эмоциональным выражением в высказывании, который извлекается из известной базы здравого смысла ATOMIC [3]. В результате Weighted-F1 score модели в наборе данных MELD достигла 66.10%, что демонстрирует превосходную производительность CISPER по сравнению с современными моделями ERC, а также эффективность использования контекстной информации и здравого смысла для повышения производительности.

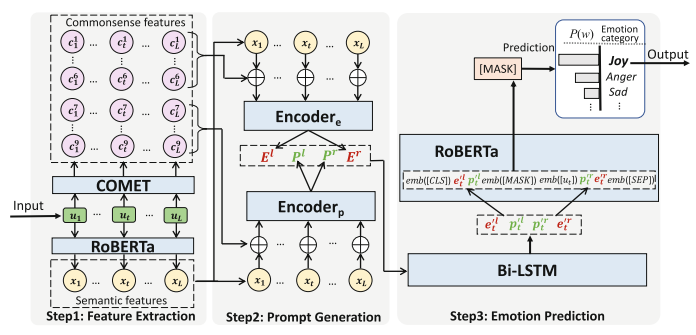


рис. 1 - Архитектура CISPER

## 2.2 CoMPM: Context Modeling with Speaker’s Pre-trained Memory Tracking for Emotion Recognition in Conversation

В работе «CoMPM: Context Modeling with Speaker’s Pre-trained Memory Tracking for Emotion Recognition in Conversation» [4] авторы (Lee J и т.д.) извлекают предварительно обученную память, используя предварительно обученную языковой модели в качестве экстрактора внешних знаний. Предлагаемая модель CoMPM состоит из модуля Context Embedding (CoM) и модуля предварительно обученной памяти (PM). На рисунке 2 показана архитектура CoMPM. CoM улавливает основное влияние всех предыдущих высказываний на эмоции текущего говорящего. Поэтому авторы предложили контекстную модель для обработки связи между текущим и предыдущим высказываниями. PM использует только предыдущие высказывания говорящего, чтобы отразить знания говорящего. В результате Weighted-F1 score модели для эмоций и тональности в наборе данных MELD достигли 66.52% и 73.08% соответственно. Результаты экспериментов показывают, что каждый модуль эффективен для повышения производительности. CoMPM превосходит базовые показатели как для двухсторонних, так и для многосторонних наборов данных и достигает современного уровня среди систем, которые не используют внешние знания.

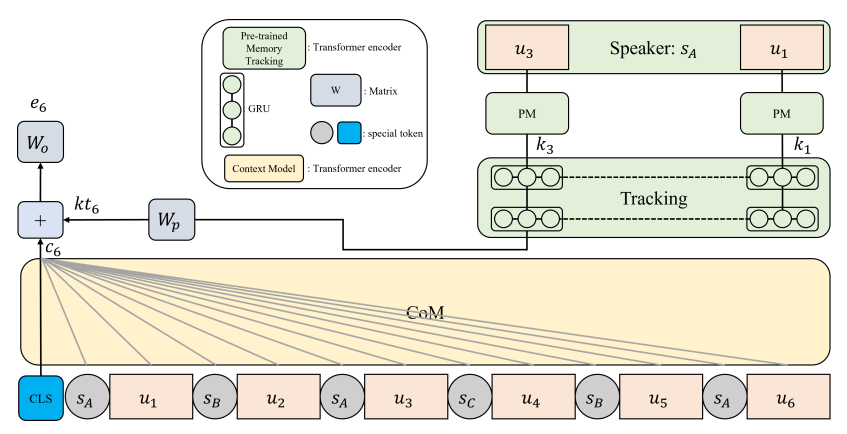


рис. 2 - Архитектура CoMPM

## 2.3 DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation

В статье Ghosal D и т.д. «DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation» [5] представлется сверточная сеть Dialogue Graph (DialogueGCN), которая моделирует зависимости между говорящими и самозависимость для улучшения понимания контекста для обнаружения эмоций на уровне высказывания в разговорах. Для извлечения текстовых признаков из транскрипции высказываний была использована сверточная нейронная сеть на основе word2vec. По трем эталонным наборам данных ERC DialogueGCN значительно превосходит существующий уровень техники.

## 2.4 Supervised Prototypical Contrastive Learning for Emotion Recognition in Conversation

В работе «Supervised Prototypical Contrastive Learning for Emotion Recognition in Conversation» [6] авторы (Song X и т.д.) предложили потерю контролируемого прототипного контрастного обучения (SPCL) для задачи ERC. Используя прототипную сеть, SPCL нацелен на решение проблемы несбалансированной классификации с помощью контрастного обучения и не требует большого размера пакета. Экспериментальные результаты демонстрируют эффективность предлагаемой нами стратегии обучения учебной программе. SimCSE [7], предварительно обученная языковая модель, используется в качестве базовой модели. Сочетая функция потери SPCL и стратегию curriculum learning, наконец, позволяет достичь самых современных результатов по трем широко используемым эталонным тестам. В результате Weighted-F1 score модели в наборе данных MELD достигла 67.23%.

## 2.5 Multimodal Representations Learning Based on Mutual Information Maximization and Minimization and Identity Embedding for Multimodal Sentiment Analysis

В статье Zheng J и т.д. «Multimodal Representations Learning Based on Mutual Information Maximization and Minimization and Identity Embedding for Multimodal Sentiment Analysis» [8] предлагается мультимодальная модель представления, основанная на максимизации и минимизации взаимной информации и внедрении идентичности (MMMIE). Для контекстной информации, вдохновленной position embedding, предложенным в Transformer, авторы предложили Identity Embedding и добавили его к функциям каждой модальности. Затем на основе механизма внимания и Long short-term memory (LSTM) контекстная информация может быть смоделирована на протяжении всего информационного потока. Комплексные эксперименты проводятся на двух общедоступных наборах данных, и результаты демонстрируют эффективность предложенной модели. В результате для текстовой модальности в наборе данных IEMOCAP точность и Weighted-F1 score модели достигли 61,6% и 61.0% соответственно.

## 2.6 Knowledge-Interactive Network with Sentiment Polarity Intensity-Aware Multi-Task Learning for Emotion Recognition in Conversations

В работе «Knowledge-Interactive Network with Sentiment Polarity Intensity-Aware Multi-Task Learning for Emotion Recognition in Conversations» [9] авторы (Xie Y и т.д.) предложили интерактивную сеть знаний с многозадачным обучением с учетом полярности тональности, а именно KI-Net, которая использует как знания здравого смысла, так и лексикон тональности для увеличения семантической информации. Сначала авторы применили кодировщик с учетом контекста и зависимостей для контекстного моделирования. Для дальнейшего улучшения представлений на уровне слов используются крупномасштабный граф знаний здравого смысла и лексикон тональности. Также представляется прогнозирование интенсивности полярности настроений (SPIP) на уровне фраз в качестве вспомогательной задачи. В результате micro-F1 score модели в наборе данных MELD достигла 63,24% (примерно на 5% лучше, чем KET) , что показывает, что задача SPIP обеспечивает многообещающее улучшение задачи ERC.

## 2.7 Joint learning on sentiment and emotion classification

В статье Gao W и т.д. «Joint learning on sentiment and emotion classification» [10] для оценки вероятности преобразования между двумя типами меток используется дополнительный набор данных, аннотированный метками как тональности, так и эмоций. На рисунке 3 показана архитектура данной совместной обучающей системы. В частности, на этапе обучения два отдельных помеченных набора данных используются для обучения двух отдельных классификаторов. На этапе тестирования одна выборка классифицируется обоими классификаторами и получает вероятности, относящиеся к каждой метке тональности или эмоции. Затем статистическая взаимосвязь используется для переноса меток тональности (или эмоций) на метки эмоций (или тональности). Окончательная метка образца определяется путем объединения исходной метки тональности (или эмоции) и переданной метки. Экспериментальные результаты показывают, что по сравнению с отдельными классификаторами данная совместная обучающая система может постоянно достигать лучших результатов, чем отдельные классификаторы.

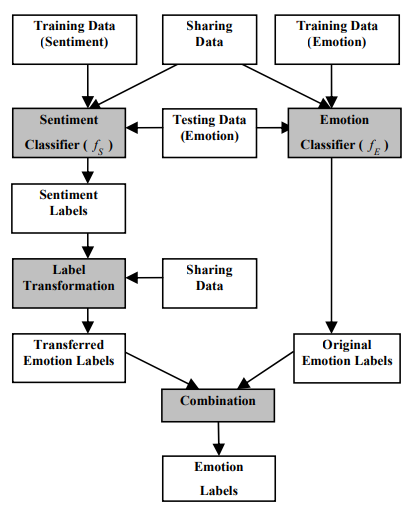


рис. 3 - Архитектура совместного обучения для классификации эмоций и тональности

## 2.8 Joint Sentiment and Emotion Classification with Integer Linear Programming

В статье Wang R и т.д. «Joint Sentiment and Emotion Classification with Integer Linear Programming» [11] предлагается основанный на целочисленном линейном программировании (ILP) подход к совместному обучению для использования связи между этими двумя задачами. В частности, авторы разработали некоторые ограничения между тональностями и эмоциями и выполнили глобальный вывод на выходных данных классификаторов тональности и эмоций с помощью ILP. В результате точности модели для классификации тональности и эмоций в наборе данных Ren-CECps corpus достигли 67.8% и 36% соответственно. Экспериментальные результаты показывают, что, по сравнению с отдельными классификаторами, данный совместный учащийся постоянно достигает лучших результатов.

## 2.9 Multimodal Emotion Recognition with Auxiliary Sentiment Information

В работе WU L и т.д. «Multimodal Emotion Recognition with Auxiliary Sentiment Information» [12] предлагается новая структура совместного обучения, которая позволяет вспомогательной задаче (мультимодальной классификации тональности) помогать основной задаче (мультимодальной классификации эмоций). Уровень частной сети используется для изучения информации о модальных внутренних эмоциях основной задачи, а вспомогательное представление эмоций основной задачи и полное представление тональности вспомогательной задачи изучаются через общий уровень. А также используется механизм внутреннего внимания для объединения информации текстовой и речевой модальности, чтобы изучить информацию о взаимодействии между модальностью. Результаты показывают, что этот метод может значительно повысить производительность задач классификации эмоций с помощью вспомогательной информации об тональности, и производительность задач классификации тональности также была в определенной степени улучшена.

## 2.10 Emotion helps Sentiment: A Multi-task Model for Sentiment and Emotion Analysis

В работе «Emotion helps Sentiment: A Multi-task Model for Sentiment and Emotion Analysis» [13] авторы (Kumar A и т.д.) авторы предложили двухуровневую многозадачную нейронную сеть, основанную на внимании, которая выполняет анализ тональности посредством анализа эмоций. Предлагаемый подход основан на двунаправленной долговременной кратковременной памяти и использует распределенный тезаурус в качестве источника внешних знаний для улучшения прогнозирования тональности и эмоций. На рисунке 4 представлена архитектура данной нейронной сети. Предлагаемая система имеет два уровня внимания для иерархического построения содержательного представления. Она была оценена на эталонном наборе данных SemEval 2016 Task 6, а также была сравнена с современными системами на Stance Sentiment Emotion Corpus. Экспериментальные результаты показывают, что предлагаемая система повышает производительность анализа тональности на 3,2 для F-score в наборе данных SemEval 2016 Task 6. Данная сеть также повышает производительность анализа эмоций на 5 для F-score в Stance Sentiment Emotion Corpus.

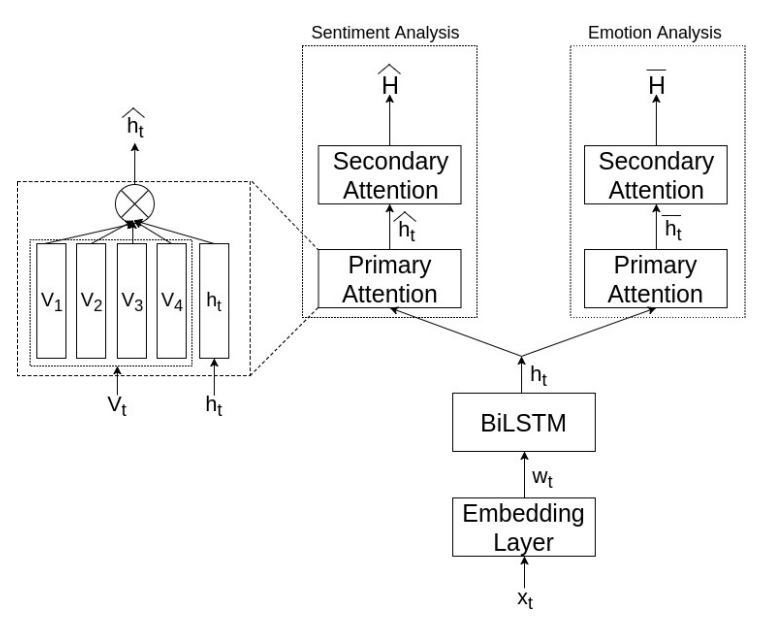


рис. 4 - Двухуровневая многозадачная сеть на основе внимания

## 2.11 Сравнительная таблица

Таблица 1. Сравнительная таблица.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | Подход | Метод извлечения признаков | Датасет | Метрики | Значения |
| 1 | CISPER | Модель RoBERTa-large для создания семантических вложений, ATOMIC | MELD | Weighted-F1 score | 66.10% |
| 2 | CoMPM  (Context Embedding Module + Pre-trained Memory Module) | RoBERTa | MELD | Weighted-F1 score | Sentiment – 73.08%  Emotion – 66.52% |
| 3 | DialogueGCN | CNN на основе Word2vec | MELD | Accuracy,  F1 score | Accuracy – 65.25%  F1 score – 64.18% |
| 4 | Функция потери SPCL, Transformer-based PLM | – | MELD | Weighted-F1 score | 67.23% |
| 5 | MMMIE: Transformer + LSTM | BERT | MELD, IEMOCAP | Accuracy,  Weighted-F1 score | Accuracy – 61.6%  Weighted-F1 score – 61.0% |
| 6 | KI-Net (Knowledge Introduction) | – | MELD | micro-F1 | 63.24% |
| 7 | Использование вероятности преобразования для переноса меток, классификатор максимальной энтропии (ME) | Bag-of-Words | Ren-CECps corpus | Accuracy | Joint\_Sentiment - 67.9%  Joint\_Emotion - 39.3% |
| 8 | совместная классификация с ILP | Bag-of-Words | Ren-CECps corpus | Accuracy | Sentiment - 67.8%  Emotion - 36% |
| 9 | Совместная мультимодальная сеть обучения эмоциям и тональностям | GloVe | MELD | Weighted-F1 score | текст - 56.43%  аудио - 41.32%  текст+аудио - 59.81% |
| 10 | Двухуровневая многозадачная сеть на основе внимания | Word2vec | SemEval 2016 Task 6, Stance Sentiment Emotion Corpus | F1-score | Emotion – 65.44%,  Sentiment – 82.10% |

## Вывод

По данной сравнительной таблице (Таблица 1) можно сделать вывод, что использование контекстной информации или введение рассуждений о здравом смысле может помочь повысить точность моделей для распознавания эмоций или тональности. Приведенные выше статьи считают, что между эмоциями и тональностями существует определенная связь, которая может повысить эффективность как моделей классификации эмоций, так и моделей классификации тональности.

# Эксполаторный анализ корпуса

Данная работа будет проведена на основе набора данных MELD, который был разработан на основе набора данных EmotionLines [14]. MELD — это мультимодальный набор данных эмоционального диалога, включающий аудио, видео и текст. MELD содержит 13 000 предложений из 1 433 диалогов сериала «Друзья», в каждом из которых более двух говорящих. Каждое предложение в диалоге было помечено любым из этих семи эмоциональных ярлыков — гнев, отвращение, печаль, радость, нейтральность, удивление и страх. MELD также имеет аннотации тональности (положительные, отрицательные и нейтральные) для каждого высказывания. По таблице 2 показана статистика набора данных MELD. В среднем в диалоге набора данных присутствуют три эмоции, и средняя продолжительность высказывания составляет 3,59 секунды.

Таблица 2. Статистика набора данных MELD. [1]

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | **Train** | **Dev** | **Test** |
| # of modality | {a,v,t} | {a,v,t} | {a,v,t} |
| # of unique words | 10643 | 2384 | 4361 |
| Avg. utterance length | 8.03 | 7.99 | 8.28 |
| Max. utterance length | 69 | 37 | 45 |
| Avg. # of emotions per dialogue | 3.30 | 3.35 | 3.24 |
| # of dialogues | 1039 | 114 | 280 |
| # of utterances | 9989 | 1109 | 2610 |
| # of speakers | 260 | 47 | 100 |
| # of emotion shift | 4003 | 427 | 1003 |
| Avg. duration of an utterance | 3.59s | 3.59s | 3.58s |

По таблицам 3 и 4 и рисункам 5 можно отметить, что распределение классов в MELD несбалансированное, и «нейтральная» преобладают в корпусе.

Таблица 3. Распределение эмоций в наборах данных MELD. [1]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Train** | **Dev** | **Test** | **Total** |
| Anger | 1109 | 153 | 345 | 1607 |
| Disgust | 271 | 22 | 68 | 361 |
| Fear | 268 | 40 | 50 | 358 |
| Joy | 1743 | 163 | 402 | 2308 |
| Neutral | 4710 | 470 | 1256 | 6436 |
| Sadness | 683 | 111 | 208 | 1002 |
| Surprise | 1205 | 150 | 281 | 1636 |

Таблица 4. Распределение тональности в наборах данных MELD. [1]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Train** | **Dev** | **Test** | **Total** |
| Negative | 2945 | 406 | 833 | 4184 |
| Neutral | 4710 | 470 | 1256 | 6436 |
| Positive | 2334 | 233 | 521 | 3088 |

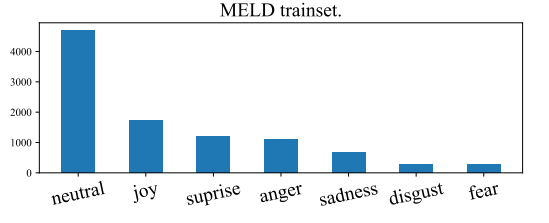


рис. 5 - Распределение эмоций в обучающих наборах данных MELD

На рисунке 6 a) и b) представлены распределения эмоций и тональности между актерами по MELD соответственно, а c) показано распределение диалогов в корпусе между актерами.

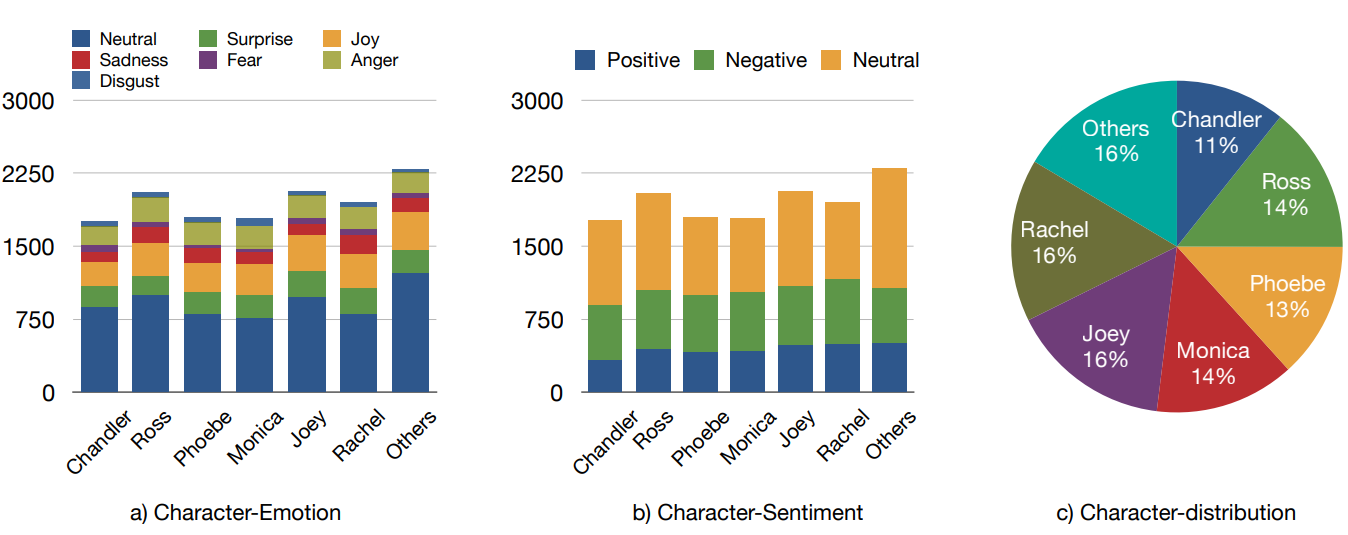


рис. 6 - Распределение эмоций и тональности между актерами по MELD [1]

**Вывод:**

* В среднем в диалоге набора данных присутствуют три эмоции. Смещение эмоций говорящего в диалоге делает задачу распознавания эмоций очень сложной.
* Распределение классов в MELD несбалансированное, и «нейтральная» преобладают в корпусе.
* Наборов данных MELD содержит многосторонние диалоги, которые более сложны по сравнению с диадическими.

# Baseline

В данной части курсовой работы были использованы два метода извлечения признаков: Word2Vec и FastText. Оба этих метода представляют собой алгоритмы глубокого обучения, которые используются для получения векторных представлений слов. Word2Vec — это алгоритм машинного обучения, используемый для обработки естественного языка, в частности, для создания векторных представлений слов. Он представляет слова в виде векторов в пространстве высокой размерности, где схожие слова находятся ближе друг к другу, а несхожие слова - далеко друг от друга. Word2Vec может быть обучен с использованием моделей непрерывного мешка слов (CBOW) или skip-gram. На рисунке 7 и 8 представлен код для обучения модели Word2Vec самостоятельно и загрузки предобученной модели word2vec-GoogleNews-vectors [15].

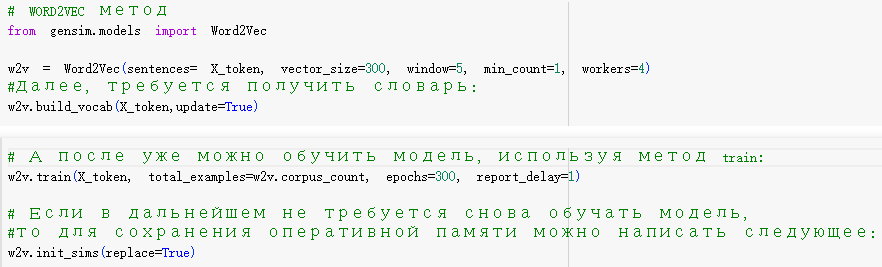


рис. 7 – код для обучения модели Word2Vec самостоятельно

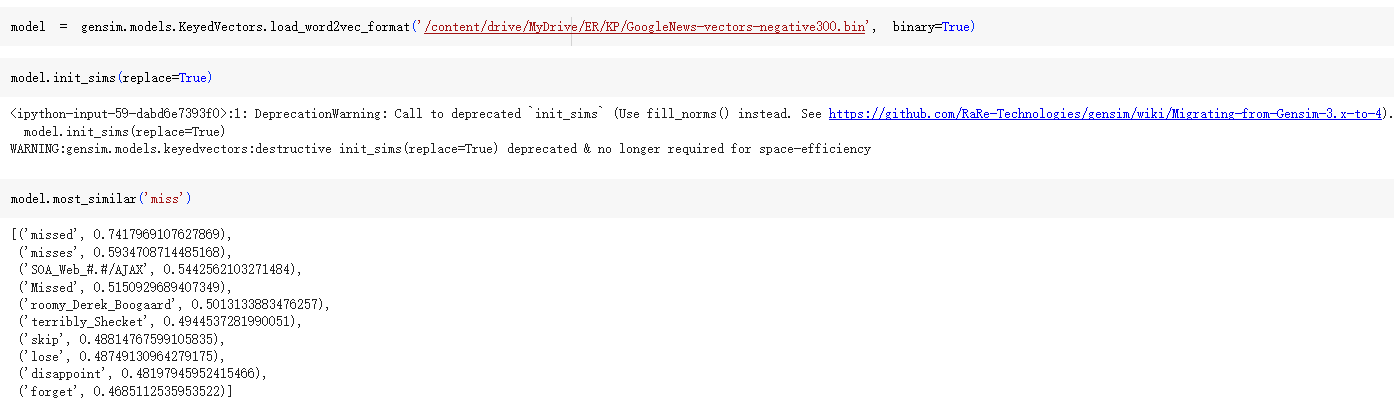


рис. 8 – код для загрузки предобученных Word2vec-признаков

Word2vec является одним из наиболее распространенных методов векторизации текста, основанным на предсказании контекстных слов для каждого слова в корпусе. В свою очередь, FastText является расширением Word2vec, которое учитывает морфологическую информацию слова. FastText (Fast Text) — это алгоритм машинного обучения для обработки естественного языка, разработанный компанией Facebook. Он похож на Word2Vec, но в отличие от него учитывает не только отдельные слова, но и их подслова (n-граммы), что позволяет строить представления для слов с неизвестными или редкими формами. FastText может использоваться для классификации текстов, кластеризации слов, поиска ближайших соседей и других задач. В этапе извлечения признаков FastText была обучена модель FastText самостоятельно (рис. 9), а также и была загружена предобученные признаки (рис. 10) для английского языка, обученные в Common Crawl и Википедии с использованием fastText [16].

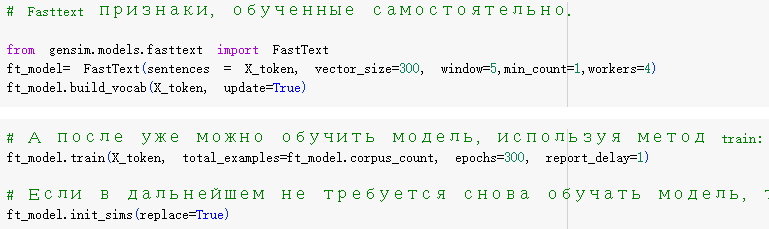


рис. 9 – код для обучения модели FastText самостоятельно



рис. 10 – код для загрузки предобученных FastText-признаков

Далее, использованы несколько традиционных классификаторов, таких как SVM, Дерево решений, Случайный лес и К ближайших соседей.

* SVM (Support Vector Machine): Линейный классификатор, который строит разделяющую гиперплоскость между классами и минимизирует ошибки классификации.
* Дерево решений (Decision Tree): Классификатор, который строит дерево решений, где каждый узел представляет собой условие на признаки, а листья - классы.
* Случайный лес (Random Forest): Ансамбль из нескольких деревьев решений, которые обучаются на разных подмножествах признаков и объектов.
* К ближайших соседей (KNN - K Nearest Neighbors): Классификатор, который основывается на близости между объектами в пространстве признаков. Для классификации нового объекта вычисляется расстояние до ближайших k объектов из обучающей выборки, и классификация происходит по большинству голосов.

Для каждого из классификаторов были произведены соответствующие эксперименты, на основе которых были получены результаты (рис. 11 - 18). Для оценки результатов классификации были использованы метрики F1-score и weighted avg f1-score. F1-score — это среднее гармоническое между precision и recall. Она позволяет оценить не только точность, но и полноту классификации и учитывает баланс между ними. Вычисляется для каждого класса отдельно и потом усредняется. Weighted avg f1-score — это взвешенное среднее по всем классам F1-score. Вес каждого класса определяется его долей в общем числе объектов. Обе метрики позволяют оценить качество классификации и учитывать несбалансированность классов. F1-score подходит для задач с равными важностями precision и recall, а weighted avg f1-score используется в задачах, где классы имеют разную важность или объем.

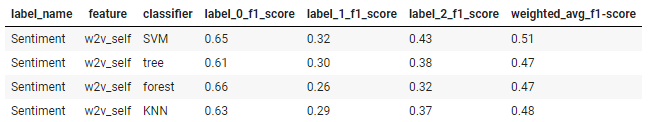


рис. 11 – результаты классификаторов для тональности (Word2vec-признаки, обученные самостоятельно)

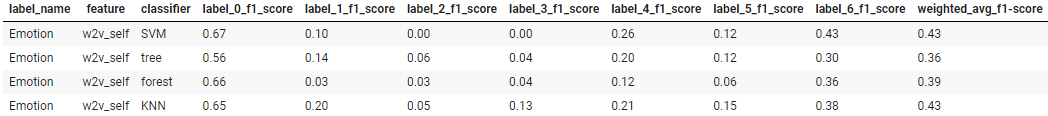


рис. 12 – результаты классификаторов для эмоций (Word2vec-признаки, обученные самостоятельно)

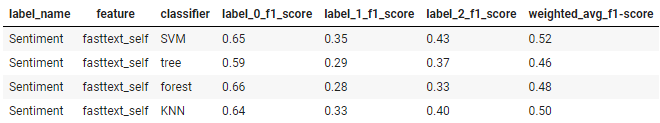


рис. 13 – результаты классификаторов для тональности (Fasttext-признаки, обученные самостоятельно)

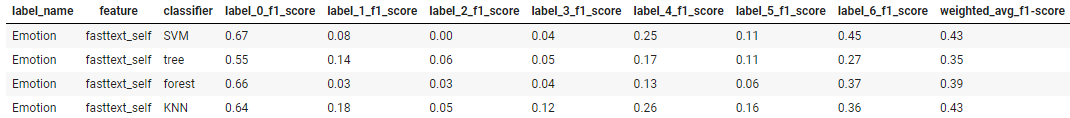


рис. 14 – результаты классификаторов для эмоций (Fasttext-признаки, обученные самостоятельно)

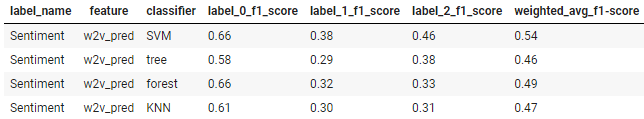


рис. 15 – результаты классификаторов для тональности (Предобученные Word2vec-признаки)

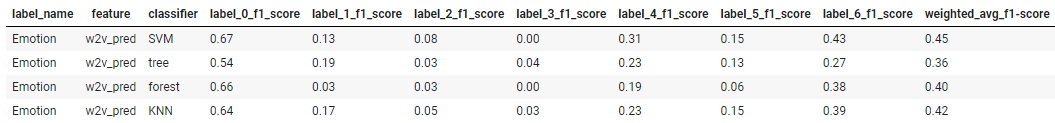


рис. 16 – результаты классификаторов для эмоций (Предобученные Word2vec-признаки)

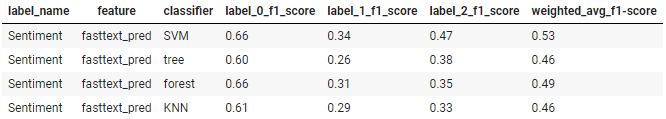


рис. 17 – результаты классификаторов для тональности (Предобученные Fasttext -признаки)

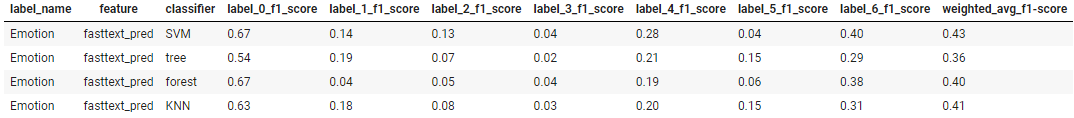


рис. 18 – результаты классификаторов для эмоций (Предобученные Fasttext -признаки)

# Усложнение моделей

Для анализа эмоций и тональности на основе текстов MELD было опробовано несколько различных типов нейронных сетей, включая нейронные сети прямого распространения, сверточные нейронные сети, рекуррентные нейронные сети (RNN), сети LSTM и управляемые рекуррентные блоки (GRU). Эти модели имеют различные слои, функции активации, dropout и другие гиперпараметры. Цель использования различных optimizer и scheduler в экспериментах - найти наилучшую комбинацию гиперпараметров для улучшения производительности моделей. Adam, RMSprop и SGD являются распространенными алгоритмами оптимизации с различными характеристиками в процессе градиентного спуска функции потерь. Методы расписания изменения темпа обучения, такие как ExponentialLR, MultistepLR и Step, могут эффективно контролировать скорость и точность обучения модели. В этом эксперименте было обнаружено, что использование RMSprop в качестве optimizer и MultistepLR в качестве scheduler может значительно улучшить точность модели.

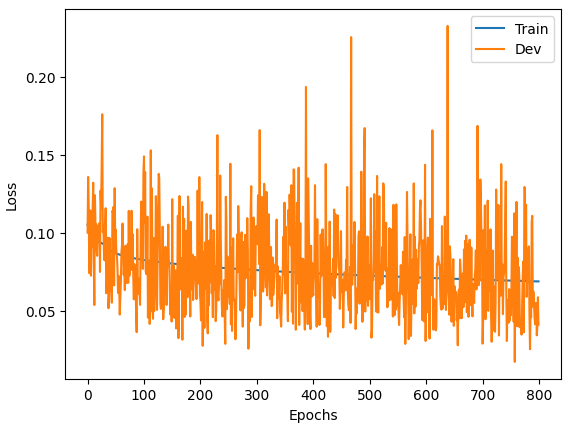


рис. 19 – пример графика кривых функции потери

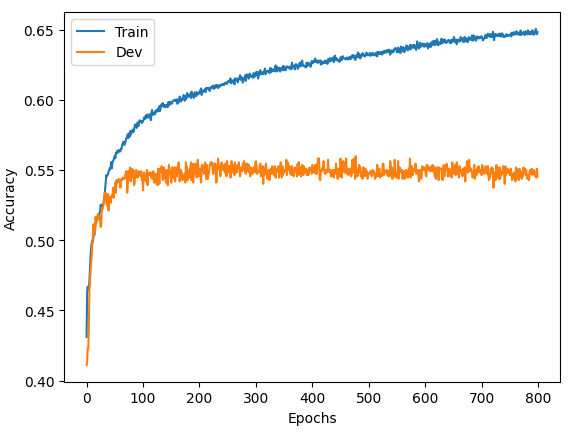


рис. 20 – пример графика кривых точности

Чтобы решить проблему сильного несбалансированности набора данных, в данной работе была использована функция потерь Focal Loss, которая назначает больший вес неправильно классифицированным примерам из миноритарного класса, что заставляет модель больше обращать внимание на эти сложные примеры. Формула функции Focal Loss имеет следующий вид:

FL(p\_t) = -alpha\_t \* (1 - p\_t)^gamma \* log(p\_t)

где,

* p\_t - предсказанная вероятность принадлежности к правильному классу. alpha\_t - вес ошибочно классифицированных примеров из менее представительного класса. Он вычисляется как 1 / количества примеров в классе или может быть задан вручную.
* gamma - фактор, который позволяет настроить важность ошибочно классифицированных примеров. Когда gamma равен 0, формула эквивалентна кросс-энтропийной функции потерь. Если значение gamma больше 0, то функция Focal Loss будет больше штрафовать ошибочно классифицированные примеры, а если значение gamma меньше 0, то функция будет менее штрафовать такие примеры.
* log - натуральный логарифм.

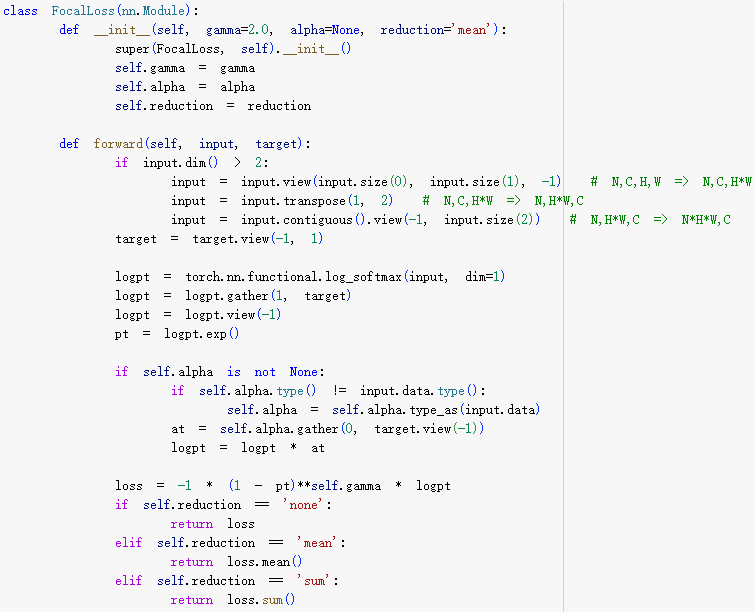


рис. 21 – Определение функции Focal Loss

В процессе экспериментов настраивала гиперпараметры, добавляла или удаляла слои, использовала различные функции активации и dropout, чтобы избежать проблем переобучения и недообучения. Ниже представлены результаты нескольких успешно обученных моделей в таблице.

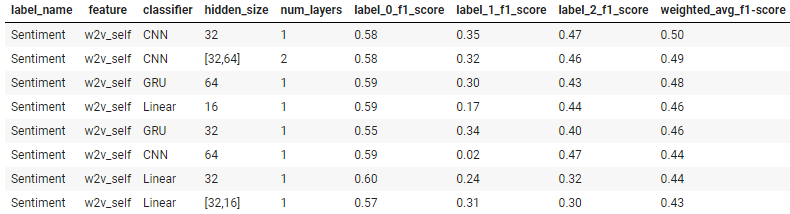


рис. 22 – результаты классификаторов для тональности (Word2vec-признаки, обученные самостоятельно)

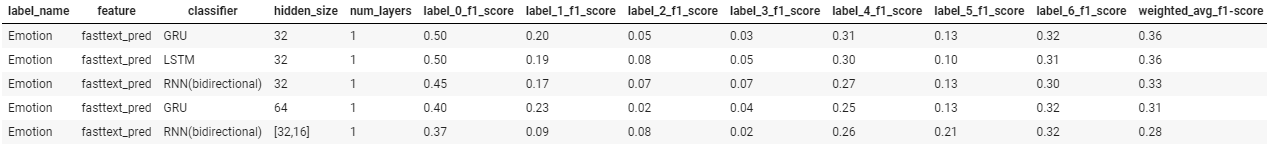


рис. 23 – результаты классификаторов для эмоций (Word2vec-признаки, обученные самостоятельно)

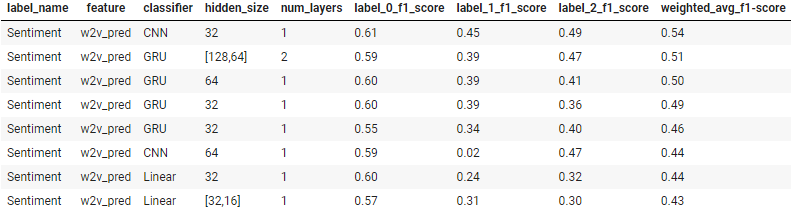


рис. 24 – результаты классификаторов для тональности (Предобученные Word2vec-признаки)

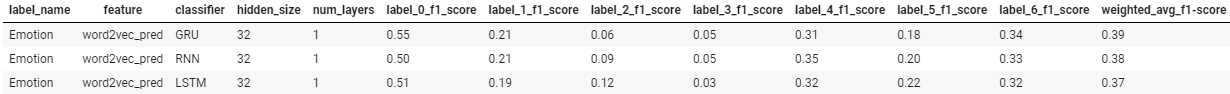


рис. 25 – результаты классификаторов для эмоций (Предобученные Word2vec-признаки)

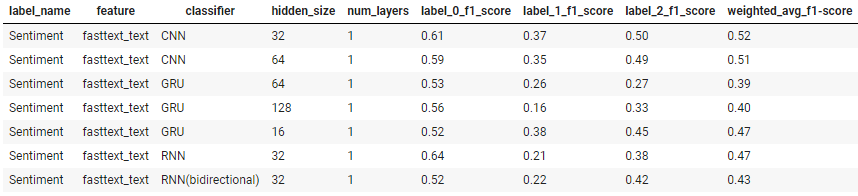


рис. 26 – результаты классификаторов для тональности (Fasttext-признаки, обученные самостоятельно)

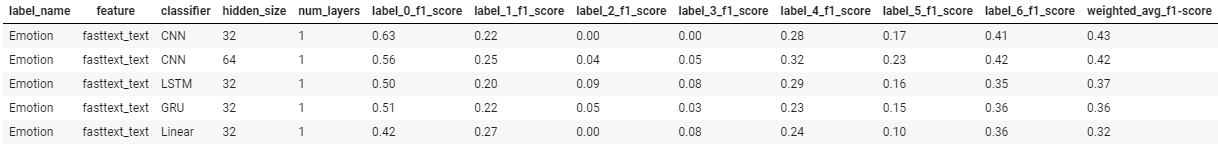


рис. 27 – результаты классификаторов для эмоций (Fasttext-признаки, обученные самостоятельно)

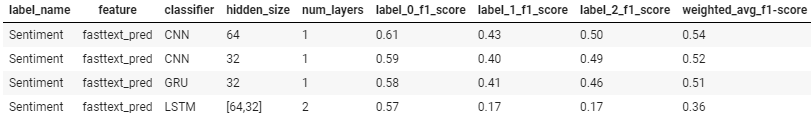


рис. 28 – результаты классификаторов для тональности (Предобученные Fasttext -признаки)

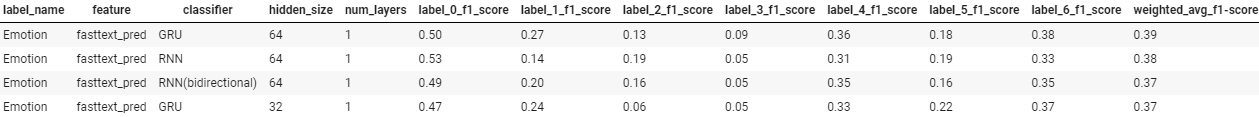


рис. 29 – результаты классификаторов для эмоций (Предобученные Fasttext -признаки)

Для дальнейшего улучшения точности модели была предложена следующая архитектура (как показано на рисунке 30). Данная архитектура включает классификатор тональности и классификаторы позитивных и негативных эмоций. Сначала извлекается с помощью методов Word2vec или fastText признаки текста, которые затем подаются на вход классификатору для тональности. Если классификатор тональности определяет тональность текста как нейтральную, тогда эмоция текста также считается нейтральной. Если тональность текста является положительной, то данные передаются в классификатор положительных эмоций. В обратном случае, если тональность отрицательная, данные передаются в классификатор отрицательных эмоций. Положительные эмоции включают радость и удивление, а отрицательные - гнев, отвращение, страх, грусть и удивление.

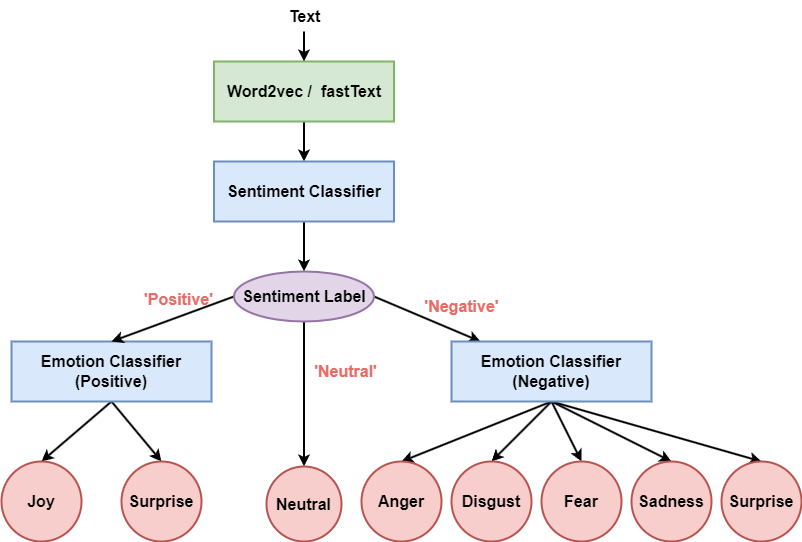


рис. 30 – Архитектура сети

На рисунках 31 – 33 показаны коды для реализации, обучения и тестирования данной архитектуры сети.

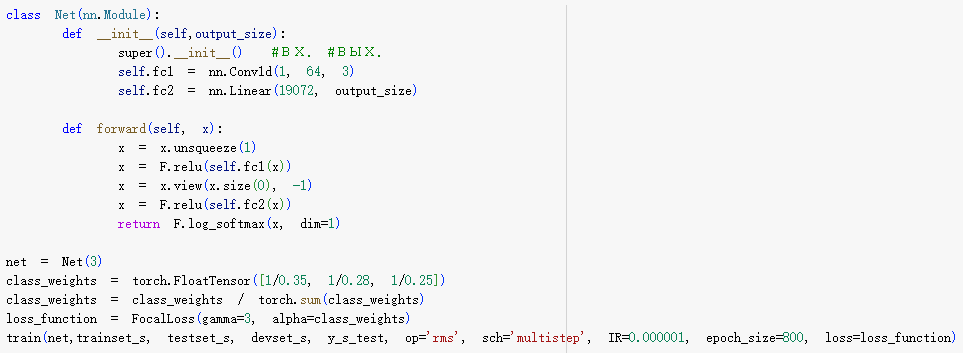


рис. 31 – обучение классификатора для тональности

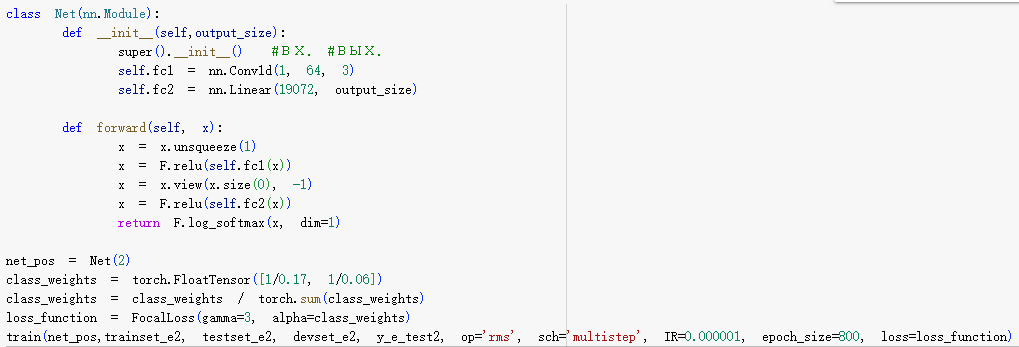


рис. 32 – обучение классификатора для позитивных эмоций

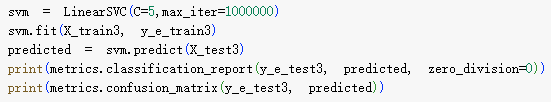


рис. 33 – обучение классификатора для негативных эмоций

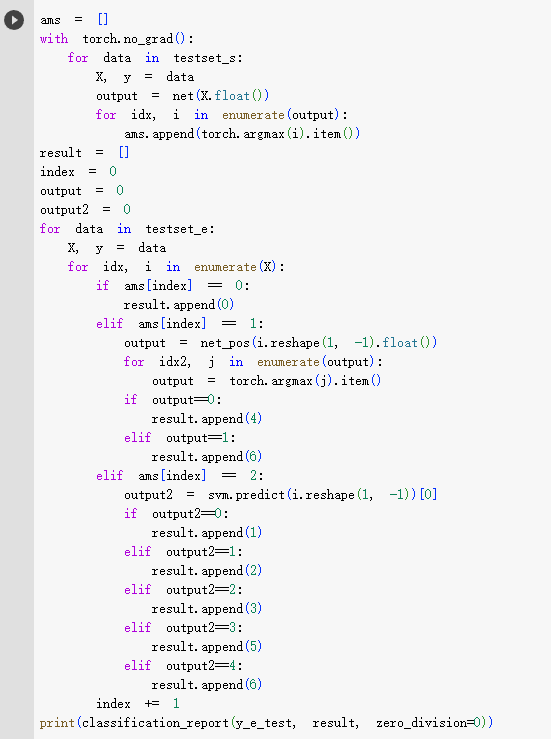


рис. 34 – код для тестирования данной модели

Для оценивания модели используется взвешенное среднее значение F1-меры, которая учитывает не только точность (precision) и полноту (recall) каждого класса, но также и их дисбаланс при подсчете среднего значения. В результате окончательное взвешенное среднее значение F1-меры составляет 42%.

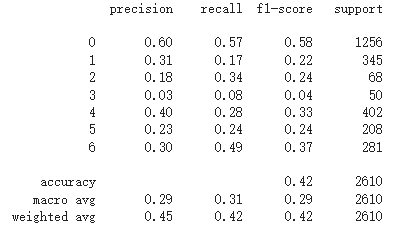


рис. 35 – результат тестирования данной модели

**Вывод**

В результате окончательное взвешенное среднее значение F1-меры составляет 42%. Взвешенное среднее значение F1-меры для класса "отвращение" составляет 24%.

# Заключение

В результате данной работы были изучены различные подходы к извлечению признаков текста и нейронные сети, часто используемые для классификации эмоций и тональности.

Была разработана и обучена модель на основе предложенной архитектуры для совместного анализа тональности и эмоций. На тестовой выборке обученная модель показала взвешенное среднее значение F1-меры не менее 42% и улучшение F1-меры для класса "отвращение". Эти результаты свидетельствуют о применимости предложенной архитектуры для задач совместного анализа тональности и эмоций.

Дальнейшее развитие данной работы может быть направлено на повышение точности каждого классификатора в архитектуре, а также на оптимизацию гиперпараметров и использование дополнительных признаков.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. S. Poria, D. Hazarika, N. Majumder, G. Naik, R. Mihalcea, E. Cambria. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversation. (2018)
2. Yi J, Yang D, Yuan S, et al. Contextual Information and Commonsense Based Prompt for Emotion Recognition in Conversation[C]//Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part II. Cham: Springer International Publishing, 2023: 707-723.
3. Sap, M., Le Bras, R., et al.: Atomic: an atlas of machine commonsense for if-then reasoning. In: Proceedings of AAAI (2019)
4. Lee J, Lee W. CoMPM: Context Modeling with Speaker's Pre-trained Memory Tracking for Emotion Recognition in Conversation[J]. arXiv preprint arXiv:2108.11626, 2021.
5. Ghosal D, Majumder N, Poria S, et al. Dialoguegcn: A graph convolutional neural network for emotion recognition in conversation[J]. arXiv preprint arXiv:1908.11540, 2019.
6. Song X, Huang L, Xue H, et al. Supervised prototypical contrastive learning for emotion recognition in conversation[J]. arXiv preprint arXiv:2210.08713, 2022.
7. Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In Empirical Methods in Natural Language Processing (EMNLP).
8. Zheng J, Zhang S, Wang X, et al. Multimodal Representations Learning Based on Mutual Information Maximization and Minimization and Identity Embedding for Multimodal Sentiment Analysis[J]. arXiv preprint arXiv:2201.03969, 2022.
9. Xie Y, Yang K, Sun C J, et al. Knowledge-interactive network with sentiment polarity intensity-aware multi-task learning for emotion recognition in conversations[C]//Findings of the Association for Computational Linguistics: EMNLP 2021. 2021: 2879-2889.
10. Gao W, Li S, Lee S Y M, et al. Joint learning on sentiment and emotion classification[C]//Proceedings of the 22nd ACM international conference on Information & Knowledge Management. 2013: 1505-1508.
11. Wang R, Li S, Zhou G, et al. Joint sentiment and emotion classification with integer linear programming[C]//Database Systems for Advanced Applications: DASFAA 2015 International Workshops, SeCoP, BDMS, and Posters, Hanoi, Vietnam, April 20-23, 2015, Revised Selected Papers 20. Springer International Publishing, 2015: 259-265.
12. Wu L, Liu Q, Zhang D, et al. Multimodal Emotion Recognition with Auxiliary Sentiment Information[J]. Beijing Da Xue Xue Bao, 2020, 56(1): 75-81.
13. Kumar A, Ekbal A, Kawahra D, et al. Emotion helps sentiment: A multi-task model for sentiment and emotion analysis[C]//2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019: 1-8.
14. Chen, S.Y., Hsu, C.C., Kuo, C.C. and Ku, L.W. EmotionLines: An Emotion Corpus of Multi-Party Conversations. arXiv preprint arXiv:1802.08379 (2018).
15. Модель word2vec-GoogleNews-vectors. URL: <https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/edit?resourcekey=0-wjGZdNAUop6WykTtMip30g>
16. Предобученные признаки для английского языка, обученные в Common Crawl и Википедии с использованием fastText. URL: https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.en.300.bin.gz