UKRAINIAN CATHOLIC UNIVERSITY

MASTER THESIS

Generation of sport news articles from match text commentary

Author:
Denys PORPLENKO

Supervisor:
PhD. Valentin MALYKH

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in the

Department of Computer Sciences
Faculty of Applied Sciences

Lviv 2020

Declaration of Authorship

I, Denys PORPLENKO, declare that this thesis titled, "Generation of sport news articles from match text commentary" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Master of Science

Generation of sport news articles from match text commentary

by Denys PORPLENKO

Abstract

Nowadays, thousands of sporting events take place every day. Most of the sports news (results of sports competitions) is written by hand, despite their pattern structure. In this work, we want to check possible or not to generate news based on the broadcast - a set of comments that describe the game in real-time. This problem solves for the Russian language and considered as a summarization problem, using extractive and abstract approaches. Among extractive models, we do not get significant results. However, we build an Oracle model that showed the best possible result equal to 0.21 F1 for ROUGE-1. For the abstraction approach, we get 0.26 F1 for the ROUGE-1 score using the NMT framework, the Bidirectional Encoder Representations from Transformers (BERT), as an encoder and text augmentation based on a thesaurus. Other types of encoders do not show significant improvements.

Acknowledgements

First of all, I want to thank my supervisor Valentin Malykh from the Moscow Institute of Physics and Technology, who gave me much useful advice regarding the experiments, results, structure, and content of this work. I am also grateful to Vladislav Denisov from Moscow Technical University of Communications and Informatics for providing data corpus and Open Data Science community [1] who was granting computational resources for research.
Also, I am grateful to the supervisor of seminars Hanna Pylieva (Ukrainian Catholic University, Data Science Master program alumnus) for discussing progress, past and future stages of this project.
Finally, I want to thank Ukrainian Catholic University and Oleksii Molchanovskyi personally for the most progressive master program in data science in Ukraine and for the opportunity to take part in it.

[1] https://ods.ai/

Contents

Declaration of Authorship **ii**

Abstract **iii**

Acknowledgements **iv**

1 Introduction **1**
1.1 Motivation 1
1.2 Goals of the master thesis 2
1.3 Thesis structure 2

2 Related work **3**
2.1 Extractive and abstarctive approaches for summarization 3
2.2 Generating news and sport summaries 4

3 Dataset description **6**
3.1 Broadcast 6
3.2 News 7
3.3 Additional information about dataset 10

4 Background and theory information **12**
4.1 Metrics 12
4.1.1 ROUGE 12
4.1.2 Cross-entropy 13
4.2 Extractive approach 14
4.2.1 PageRank 14
4.2.2 TextRank 14
4.2.3 LexRank 15
4.3 Word embedding 16
4.3.1 Word2vec 16
4.3.2 FastText 16
4.4 Abstractive approach 17
4.4.1 RNN 17
4.4.2 LSTM 19
4.4.3 Sequence-to-sequence model 19
4.4.4 Neural Machine Translation 20
4.4.5 Attention mechanism 20
4.4.6 Transformers 21
4.4.7 BERT 22
4.5 Differences in broadcasts styles and news 23

5 Model **25**
5.1 OpenNMT 25
5.2 PreSumm 25

6 Experiments **28**
6.1 TextRank approaches 28
Implementation details 28
Results and examples 29
6.2 LexRank 29
Implementation details 29
Results and examples 30
6.3 Oracle 31
Implementation details 31
Results and examples 31
6.4 OpenNMT 32
Implementation details 32
6.5 PreSumm 33
Implementation details 33
6.5.1 BertSumAbs 35
6.5.2 RuBertSumAbs 36
6.5.3 BertSumExtAbs 36
6.5.4 BertSumAbs1024 37
6.5.5 OracleA 37
6.5.6 BertSumAbsClean 38
6.5.7 AugAbs 38
6.6 Human evaluation 40
6.7 Compare human judgment and score ROUGE 41

7 Conclusion **43**
7.1 Contribution 43
7.2 Future work 44

A Examples of generated news **45**

Bibliography **58**

List of Figures

3.1 Number of comments per each sport game. 8
3.2 Number of tokens (splited by space) per one broadcast. 8
3.3 Number of news per one sport games. News splited for two group: before and after news. 10
3.4 Number of tokens (splited by space) per one sport game. We cut news with more than 1000 tokens. 11

4.1 Stages of TextRank algorithm. 15
4.2 Two-dimensional PCA projection of countries and their capital cities. The figure illustrates the ability of the model to learn implicitly the relationships between them without any supervised information about what a capital city means. Source: (Mikolov et al., 2013a) 17
4.3 The CBOW architecture predicts the current word based on the context, and the Skip-gram predicts surrounding words given the current word. Source: (Mikolov et al., 2013b) 18
4.4 A Recurrent Neural Network(RNN). Three time-steps are shown. . . . 18
4.5 The detailed internals of a LSTM. Source: (*CS224n: NLP with Deep Learning*). 20
4.6 The graphical illustration of the proposed model trying to generate the t-th target word yt given a source sentence $(x_1, x_2, ..., x_T)$. Source: (Bahdanau, Cho, and Bengio, 2014). 21
4.7 The Transformer - model architecture. Source: (Vaswani et al., 2017). . 22
4.8 The figure shows pre-training and fine-tuning procedures for the BERT model. The same architectures and model parameters (to initialize models for different down-stream tasks) are used in both pre-training and fine-tuning stages. During fine-tuning, all parameters are fine-tuned. Source: (Devlin et al., 2018). 23

5.1 Architecture of the original BERT model (left) and BERTSUM (right). Source: Liu and Lapata, 2019 26

6.1 This figure shows the distributation on number of sentences using Oracle model. 31
6.2 This figure shows the distributation number of tokens in news (target for Presumm experiments). 34
6.3 This figure shows the distributation number of tokens in broadcasts (source for Presumm experiments). 35

List of Tables

3.1 Examples of comments for a same sport game. 7
3.2 Examples of news for sport games 9
3.4 Examples of service/advertising information inside broadcasts and news. 10
3.3 Examples of after sport game news, with before sport game or general contexts 11

6.1 ROUGE scores of TextRank approaches.PageRank W2V and PageRank FT models based on PageRank and used word2vec and FastText models. TextRank Gns - an algorithm from the Gensim toolkit. 29
6.2 ROUGE-1/ROUGE-2/ROUGE-L score of LexRank results. 30
6.3 Examples of summaries generated using LexRank approach. 30
6.4 ROUGE scores for all exptractive approaches. 31
6.5 ROUGE scores for summaries generated by using model BertSumAbs. 35
6.6 ROUGE scores for summaries generated by using RuBertSumAbs. . . 36
6.7 ROUGE scores for summaries generated by using BertSumExtAbs. . . 36
6.8 ROUGE scores for summaries generated by using BertSumAbs1024. . 37
6.9 ROUGE scores for summaries generated using two oracle-based models. 38
6.10 ROUGE scores for summaries generated from BertSumAbsClean model. 38
6.11 ROUGE scores from all models. 39
6.12 Comparison between human annotators and ROUGE scores for 5 news along different dimensions and models. HJS - human judgment score. 42

A.1 Examples of summaries generated using BERTABS model. 45
A.9 Table show components of test example 1 for the human evaluation. . 47
A.10 Table show components of test example 2 for the human evaluation. This example includes broadcast information, using as the input source, candidate news, generated by the model, and "gold" news, written by a human. 49
A.2 Examples of summaries generated using BertSumOracletAbs and BertSumOracleExtAbs models. They showed similar results, so we do not divide samples of summaries by models. 50
A.3 Examples of summaries generated using extractive approaches: PageRank+word2vec, PageRank+fastText + Gensim TextRank. 51
A.4 Examples of summaries generated using Oracle model. 52
A.5 Examples of summaries generated using RuBERTABS model. 53
A.6 Examples of summaries generated using BertSumAbs1024 model. . . . 54
A.7 Examples of summaries generated using BertSumOracletAbs and OracleEA models. They showed similar results, so we do not divide samples of summaries by models. 55
A.8 Examples of summaries generated using AugAbsTh model. 56
A.11 Examples of summaries generated using BertSumExtAbs model. . . . 57

List of Abbreviations

NLP	**N**atural **L**anguage **P**rocessing
RNN	**R**ecurrent **N**eural **N**etwork
LSTM	**L**ong **S**hort-**T**erm **M**emory
BERT	**B**idirectional **E**ncoder **R**epresentations from **T**ransformers
NMT	**N**eural **M**achine **T**ranslation
SMT	**S**tatistical **M**achine **T**ranslation
TF-IDF	**T**erm **F**requency – **I**nverse **D**ocument **F**requency
ROUGE	**R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation
CBOW	**C**ommon **B**ag **O**f **W**ords
OOV	**O**ut-**O**f-**V**ocabulary
GRU	**G**ated **R**ecurrent **U**nit

List of Symbols

A	accuracy
P	precision
R	recall
F_1	score

Chapter 1

Introduction

1.1 Motivation

Nowadays, sports content is viral. Some events are trendy, watched by billions of people (Wikipedia contributors, 2019). Every day, thousands of events take place in the world that interest millions of people. The audience of online sports resources is quite broad (eBizMBA Inc, 2019),(SimilarWeb LTD, 2019). Even if a person watched a match, he is interested in reading the news, as there is more information in the news. Therefore, there is a great need for human resources to write this news or several for each sporting event.
Media companies have become interested in cutting costs and increasing the quality of news (Graefe, Digital Journalism, and GitBook, 2016). Some well-known publications such as the Associated Press, Forbes, The New York Times, Los Angeles Times make automatic (or "man-machine marriage" form) generating news in simple topics (routine news stories) and will also introduce research to improve the quality of such news (Graefe, Digital Journalism, and GitBook, 2016). Benefits of generating sports news:

1. The algorithm will write news faster and make the minimum number of errors.

2. The algorithm will be able to generate the same news from a different point of view, in several languages and personalizing them to an individual reader's preferences.

As we see, the above needs provide the primary motivation for our work We decided to learn how to generate news, entirely automatically, without highlighting the fit about the match, based only on textual comments of the match. Text comments are expressions, sentences that describe a game at a particular point in time. We also decided to check what results we will get using some SOTA approaches for summarization tasks (Liu and Lapata, 2019), (Klein et al., 2017a) not on CNN / Daily Mail dataset (Hermann et al., 2015a), but applied to industrial datasets (Gavrilov, Kalaidin, and Malykh, 2019) and sport.ru dataset.
In this work, we decided to generate the news and not just the short result of the match. In addition to the fact that the news can describe the score of the match, can also extend it with the details of the game ("Chelsea goalkeeper Petr Cech suffered an ankle injury in the match of the 20th..."), an interview after the event ("The coach shared his impressions of the game.."), the overall picture or situation ("Dynamo with 14 points takes sixth place in the ranks...")

1.2 Goals of the master thesis

1. Reviewing previous work on text generation using extractive and abstractive approaches.
2. Apply unsupervised techniques to calculate extractive summarization and oracle extractive model. Later apply two neural seq2seq approaches.
3. Compare the results of the study, explain why different approaches give such results. Tell about the limitations in the current task and further actions.

1.3 Thesis structure

This work is structured as follows: in chapter 2, we have an overview of related works. In chapter 4, we give a theoretical basis for the extractive and abstractive summarization approaches, the base of seq2seq models, NMT architecture, and BERT model. Chapter 5 introduces our methods and models. In Chapter 3, we describe the dataset used in current work. Also, we describe the preprocessing stage there. In Chapter 6, we present and discuss our results for each approach. Finally, in chapter 7, we will get a conclusion and set the points for further research.

Chapter 2

Related work

In current work, we try to generate news based on the comments that describe sporting games. This task belongs to problems called summarization task: generating a short conclusion from the text. At the moment, there are two approaches to solving such problems. First of it called extractive summarization, the idea of which is to choose sentences from the input text that describe the main ideas of the document as much as possible. Another one is abstractive summarization, which is to generate a text that describes the main ideas, smaller in size, and not like the input document: it can have a different style and use other words. In this chapter, we describe existing methods and studies for summarizing texts. At the end of the chapter, we will talk about works based on news generation, which summarize or describe sports events and their results.

2.1 Extractive and abstarctive approaches for summarization

The first studies in text summarization appeared back in the 1950s (A.I. Mikhailov, 1965); however, due to the limited computing resources, it was not possible to achieve good results. A new round of development began in the early 2000s, during this period we want to mention a paper (Jin and Hauptmann, 2001), in which the authors used TF-IDF algorithms for weighting and k-means to solve the title generation problem (a particular case of summarization problem). Since, at that time, there was no large corpus of data, most of the algorithms were extractive and used the unsupervised learning approaches. Here we want to note the TextRank algorithm proposed by Mihalcea and Tarau in 2004 (Mihalcea and Tarau, 2004), which showed good results; we used it in current work.

Further, the emergence of large data corpora, for example, for the Document Understanding Conference competition (Nenkova, 2005; Dang, 2006), allowed to development of extractive summarization algorithms with a teacher. We mention here the work (Wong, Wu, and Li, 2008), in which Wong and others use the support vector method to classify sentences suitable for summers. The further creation of large text corpora gave power to the development of abstractive approaches. For example, a corpus developed (Nallapati et al., 2016), which includes 700 thousand news documents in English containing brief abstracts for each paragraph. We want to note that on the one hand, this simplifies the task of summarizing the text, reducing it to summarizing a paragraph of text, but on the other hand, this approach seems to be inapplicable in practice - the summary of each paragraph, especially for a news document, may contain information that does not directly relate to the main topic of this news document.

(Rush, Chopra, and Weston, 2015) proposed an algorithm based on local attention and used the encoder-decoder approach. We also want to note that for this approach,

it is not necessary to use a text corpus in which should be an explicit indication of a part of the text (for example, a paragraph) and its summary. Later (Nallapati, Zhai, and Zhou, 2016), based on previous work, use a different type of recurrence network and obtained the state-of-the-art results. Nallapati and other authors used copying words from the input sequence to the output, thereby solving the problem with rare words. Since we have a long incoming sequence (chapter 3 describes our datasets), we want to mention another work that is based on the previous model. In this paper, Cohan and the authors proposed a summarization model for very long documents, like, science papers. They use the hierarchical encoder mechanism that models the discourse structure of a document (Cohan et al., 2018). (Cibils et al., 2018) suggested using a specialized beam search to improve the quality of the abstractive generated. Tan and co-authors suggested that a large number of errors in the automatic abstractive come from the fact that the referencing system incorrectly selects the central object in the original text. They implemented a mechanism for replacing the named entities generated by the model with the same entities from the source text, which allowed to increase the factual accuracy of the abstracts generated (Tan, Wan, and Xiao, 2017a).
Further, we would like to mention some works that use reinforcement training to solve abstractive summarization problems. Paulus, the co-author of the first work, proposed to solve the problem in two stages: In the first, the model is trained with a teacher, and in the second, its quality is improved through reinforced learning (Paulus, Xiong, and Socher, 2017). Celikyilmaz and co-authors presented a model of co-education without a teacher for two peer agents (Celikyilmaz et al., 2018).
For the Russian language, there are not many works on abstractive abstraction, which mainly appeared in the last year. First of all, this is the work of Gavrilov, Kalaidin, and Malykh (Gavrilov, Kalaidin, and Malykh, 2019), which presented a corpus of news documents, suitable for the task of generating headings in Russian. Also, in this work, was presented the Universal Transformer model as applied to the task of generating headers; this model showed the best result for Russian and English. Some other works (Sokolov, 2019; Gusev, 2019; Stepanov, 2019) was based on presented a corpus of news, which use various modifications of models based on the encoder-decoder principle.

2.2 Generating news and sport summaries

Next, we will consider works that are directly related to news generation as a summary of the text. Here we want to highlight a study by the news agency Associated Press (Graefe, Digital Journalism, and GitBook, 2016). Andreas Graefe, in this study, talks in detail about the problems, prospects, limitations, and the current state in the direction of automatic generating news. In the direction of generating the results of sports events, there is little research; here, we want to highlight two. The first is a relatively old study based on the content selection approach performed on a task-independent ontology (Bouayad-Agha, Casamayor, and Wanner, 2011; Bouayad-Agha et al., 2012; Ayala, 2019). The second that we have inspired is the graduation work of NLP Stanford 2019 student Miguel Ayala. (Ayala, 2019) used the NMT approach (Bahdanau, Cho, and Bengio, 2014) to generate a summary from the vector description (Secareanu, 2017).

At the end of the chapter, we would like to make a brief conclusion of the above. First of all, we want to note that most of the work is done for the English and Chinese text corps. These languages practically and lack a morphological change in words entirely, and also the order of words in a sentence is fixed. Thus, it is not possible to directly verify the applicability of existing summarization methods. The next, there is no single complete corpus for summarization, such as DM/NYT/XSum in the Russian language. The last is that there is no data corpus directly marked out for the current task: a corpus that contains a description of sports games and the resulting news.

Chapter 3

Dataset description

For the experiments, we used non-public data provided by sport.ru - one of the directions of which is a text broadcast of sporting events. The data provided in the form of two text entities, these are the comments from the commentator who describes the event and the news. In the provided set, there are 8781 sporting events, and each event contained several comments and news; the news was published both before and after the sporting event. A description of each entity, examples, statistical characteristic and a preprocessing process are described below.

3.1 Broadcast

The provided data consists of a set of comments for each sporting event. Table 3.1 shows examples of comments, and figure 3.1 displays the distribution number of comments for each sports events. Comments contain various types of information:

- Greetings. Ex: "Добрый день"("Hello"), "хорошего дня, любителям футбола"("have a nice day football funs");
- General information about the competition/tournament/series of games. Ex: "подниматься в середину таблицы"("rise to the middle of the table"), "пятый раз в истории сыграет в групповом турнире"("the fifth time in history will play in the group tournament");
- Information about what is happening in the game/competition. Ex: "пробил в ближний угол"("struck into the near corner"), "удар головой выше ворот"("hit (ball) head above the gate");
- Results/historical facts/plans/wishes for the players. Ex: "0:3 после сорока минут"("(score in the game) 0:3 after forty minutes"), "не забивали голов в этом сезоне"("didn't score a goal this season(players)");

Also, each comment contains additional meta-information:

- **match_id** - match identifier;
- **team1, team2** - the name of the competing teams (Real Madrid, Dynamo, Montenegro);
- **name** - the name of the league (Stanley Cup, Wimbledon. Men. Wimbledon, England, June);
- **match_time** - UNIX match time;
- **type** - an event types (nan, "yellowcard", "ball",);

Добрый день! Наш сайт поздравляет всех, кто прошедшей зимой с нетерпением считал дни до старта российской премьер-лиги. Наша первая текстовая трансляция чемпионата 2009 поможет Вам проследить за событиями, который произойдут на стадионе "Локомотив", где одноименная команда принимает гостей из "Химок". Good day! Our site congratulates everyone who counted the days before the start of the Russian Premier League. Our first text broadcast will help you follow the events that will take place at the Lokomotiv Stadium, where the team of the same name hosts guests from "Khimki".
Будем надеяться, что признаков пресловутого "весеннего" футбола на "Локомотиве" сегодня будет меньше, нежели на других аренах страны. Погода в Москве солнечная, почти тепло, да и стадион принято называть лучшим в стране, а это значит, что газон должен быть в порядке. Let's hope that there will be fewer signs of the notorious "spring" football on Lokomotiv today than on other areas in the country. The weather in Moscow is sunny, almost warm, and the stadium is usually called the best in the country, which means that the lawn should be in order.
Что ж, сегодняшних соперников можно назвать одними из самых загадочных команд сезона. Но загадочность проявляетя в них совершенно по-разному. Если болельщики хозяев верят, что кризисные времена ушли в прошлое, и команда в этом году наконец-то отважиця на чемпионский выстрел, то химкинские поклонники пока находяця в полном неведении. Well, today's competitors can be called one of the most mysterious teams of the season. But the mystery manifests itself in them in entirely different ways. If the fans of the hosts believe that crisis times are a thing of the past, and this year the team will finally dare to take the championship shot, then "Khimki" fans are still entirely in the dark.

TABLE 3.1: Examples of comments for a same sport game.

- **minute** - the minute in the sport game when the event occurred;
- **content** - a text message of the event;
- **message_time** - comment time;

We sorted by time and merged all the comments for one game into one large text and called ita broadcast. Before mearging we cleaned out unnecessary characters ("\n", "\t", "&...", html tags). There are 722067 comments in the dataset, of which we received 8781 broadcasts. In the current study, we used only text information from the commentary (field *content*). The figure 3.2 shows distribution of the number of words in each broadcast.

3.2 News

News is a text message that briefly describes the events and results of a sports game. Unlike a brief summary, news can be published before and after the match. The news that took part in the experiments contains the following information:

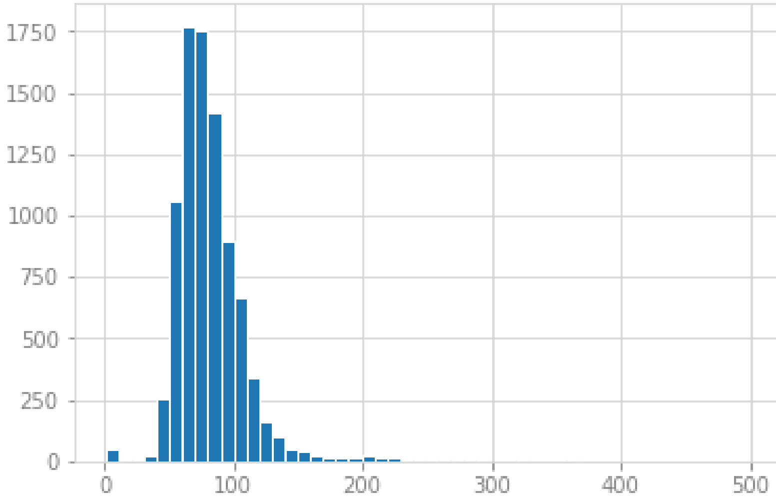

FIGURE 3.1: Number of comments per each sport game.

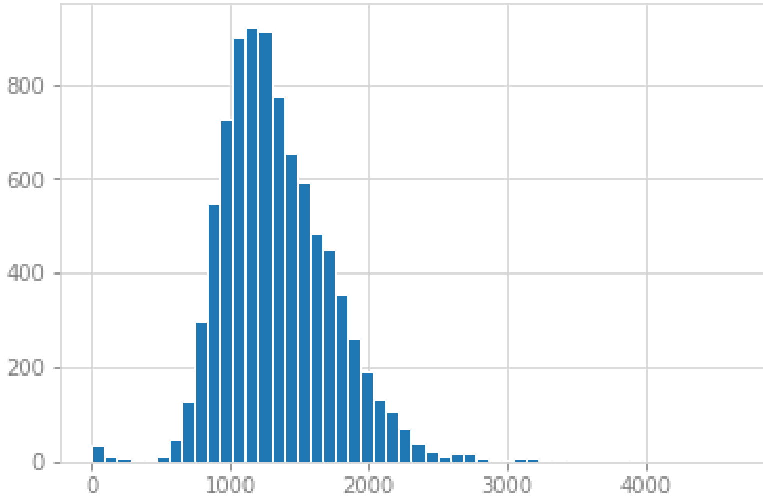

FIGURE 3.2: Number of tokens (splited by space) per one broadcast.

- Comments and interviews of a player or coach. Ex: "ребята отнеслись к матчу очень серьезно. я доволен"("The guys took the match very seriously. I am satisfied."), "мы проиграли потому, что..."("we lost because...");
- Events occurring during the game. Ex: "боковой арбитр удалил полузащитника"("side referee removes midfielder I am satisfied."), "полузащитник «Арсенала» Санти Касорла забил три мяча"(""Arsenal" midfielder Santi Cazorla scores three goals");
- General information about the competition/tournament/series of games. Ex: "сборные словакии и парагвая вышли в 1/8 финала"("national teams of Slovakia and Paraguay reached the 1/8 finals"), "выполнить задачу на турнир выйти в четверть финал"("They must complete the mission of the tournament to reach the quarter-finals.");
- Game results. Ex: "таким образом, счет стал 1:1"("thus the score was 1: 1"), "счет в серии: 0-1"("the score in the series: 0-1");

Each news contained additional meta informa-tion:

- **name** - title of news;
- **ctime** - time of news;
- **body** - text of news;
- **match_id** - sport game identificator;

Мадридский Реал выиграл в 22-м туре чемпионата Испании у Реал Сосьедада (4:1) и довел свою победную серию в домашних играх в этом сезоне до 11 матчей.Подопечные Жозе Моуриньо в нынешнем чемпионате еще не потеряли ни одного очка на Сантьяго Бернабеу.Всего победная серия Реала в родных стенах в Примере насчитывает уже 14 встреч. Соотношение мячей 45:9 в пользу королевского клуба. Real Madrid won in the 22nd round of the championship of Spain against Real Sociedad (4:1) and lead his winning streak in home games this season to 11 matches. José Mourinho's pupils in the current championship have not lost a single point in Santiago Bernabeu. The winning series of Real in the home walls in Example has already 14 meetings. Goal ratio 45: 9 in favor of the royal club.
Полузащитник питерского Зенита Анатолий Тимощук и защитник московского Динамо Денис Колодин в матче 10-го тура чемпионата России между этими командами получили по четвертой желтой карточке в этом сезоне.Таким образом, Тимощук пропустит следующий домашний матч Зенита с краснодарской Кубанью, а Колодин не примет участия в выездной игре динамовцев с Томью. Midfielder of St. Petersburg Zenith Anatoly Timoshuk and defender of Moscow Dynamo Denis Kolodin in the match of the 10th round of the championship of Russia between these teams received a fourth yellow card this season. away game of Dynamo with "Tomyu".
Полузащитник Ливерпуля Стивен Джеррард, удаленный в матче 3-го раунда Кубка Англии с МЮ (0:1), получил 6-ю красную карточку за время выступлений в составе красных.Предыдущие пять удалений капитана Ливерпуля пришлись на матчи премьер-лиги.До сегодняшнего матча с МЮ Джеррард не получал красных карточек почти пять лет. Liverpool midfielder Stephen Gerrard, removed in the FA Cup 3rd round match with Manchester United (0: 1), received the 6th red card during his appearances in the reds. The previous five removals of the captain of Liverpool have been in the Premier League matches. Before today's match, Gerrard has not received red cards with Manchester United for almost five years.

TABLE 3.2: Examples of news for sport games

The data source consists of 92997 news. We selected the news with a minimum length that was written after the match (we compared the time of the last comment from broadcast and the publish time of news). Table 3.2 shows examples of news, and figure 3.3 displays the distribution number of news per one sports game. We demonstrate the distribution of the number of words in one news in Figure 3.4.

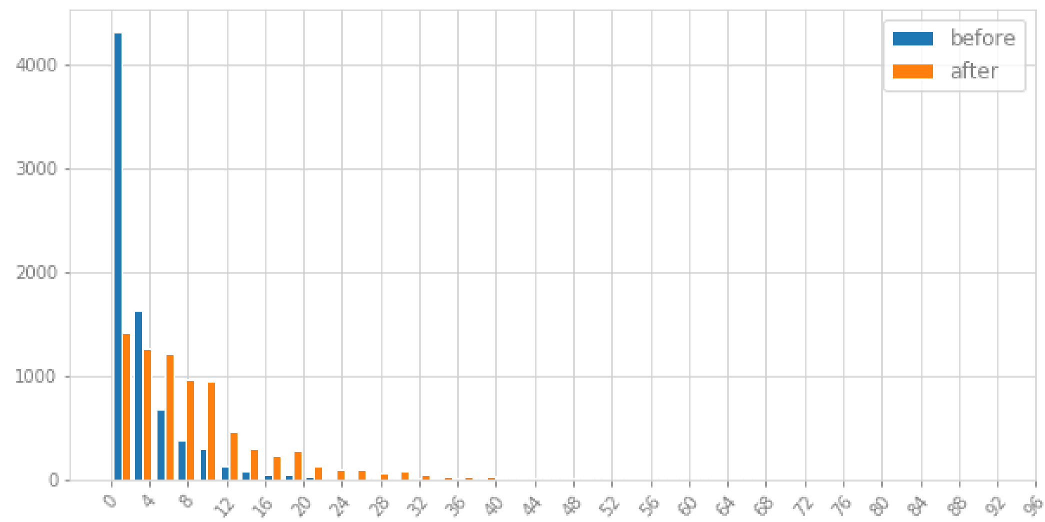

FIGURE 3.3: Number of news per one sport games. News splited for two group: before and after news.

3.3 Additional information about dataset

In our experiments described in chapter 6, we used only news written after a sporting event. We observed that although the news was published after the end of the sports event, their context was similar to the news before the match. We assume that this is an error in collecting the data. We show examples of such news in table 3.3.

We would like to emphasize that some comments and news contain advertising or general information. Table 3.4 displays examples of advertising comments. In section 6.5.6 we discuss this property of comments, how we got rid of it and what results we got.

sports.ru провел онлайн-трансляцию матча. sports.ru organized an online broadcast of the match
Таблица чемпионата Украины Статистика чемпионата Украины. Ukrainian Championship Table Ukrainian Championship Statistics
Видео приложение sports.ru для iphone и для android... Sports.ru video application for iphone and android...

TABLE 3.4: Examples of service/advertising information inside broadcasts and news.

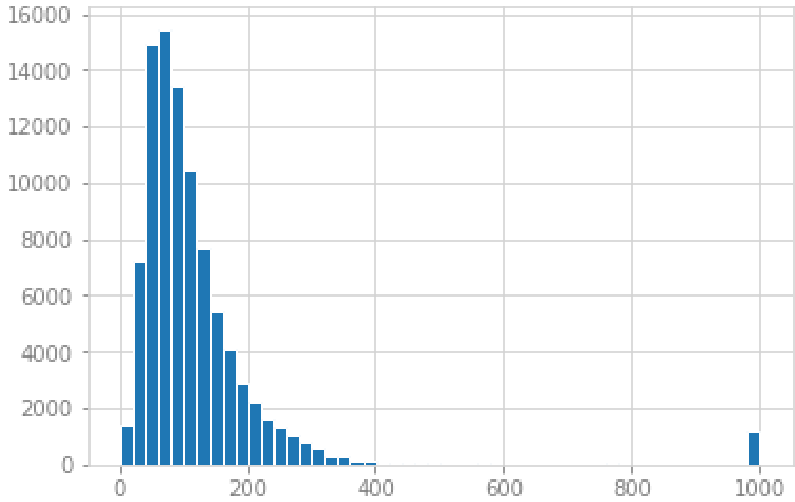

FIGURE 3.4: Number of tokens (splited by space) per one sport game. We cut news with more than 1000 tokens.

В рамках 32-го тура английской премьер-лиги «Вест Бромвич» примет «Арсенал». Предлагаем вашему внимаю стартовые составы команд.. As part of the 32nd round of the English Premier League, West Bromwich will host Arsenal. We offer you the starting list of players...
Сегодня запланировано проведение четырех матча 1/4 финала кубка Украины. «Шахтер» одолел «Карпаты» (2:1), Одесский «черноморец» в гостях переиграл «Арсенал» (2:1). «Волынь» обыграна «Днепром» (0:2). «Севастополь» справился с «Таврией» (1:1, по пенальти – 4:1). Кубок Украины1/4 финала. Today it is planned to hold four matches of the quarter-finals of the Ukrainian Cup. "Shakhtar" defeated "Karpaty" (2:1), Odessa's "Chernomorets" outplayed "Arsenal" (2:1). "Volyn" beaten by the "Dnipro" (0:2). "Sevastopol" coped with "Tavria" (1:1, on a penalty - 4:1). The cup of Ukraine 1/4 finals.
Стали известны все пары команд, которым предстоит сыграть в первом раунде плей-офф. ВОСТОК Рейнджерс (1) Оттава (8) Бостон (2) Вашингтон (7) Флорида (3) Нью-Джерси (6) Питцбург (4) Филадельфия (5) Запад Ванкувер (1) All the pairs of teams that will play in the first round of the playoffs became known. "EAST Rangers" (1), "Ottawa" (8), "Boston" (2), "Washington" (7), "Florida" (3), "New Jersey" (6), "Pittsburgh" (4), "Philadelphia" (5), "West Vancouver" (1)

TABLE 3.3: Examples of after sport game news, with before sport game or general contexts

Chapter 4

Background and theory information

In this chapter, we briefly describe the underlying technologies, approaches, algorithms that we will use for generation news from broadcasts. Models, based on these approaches were described in chapter 5, experiments and results in chapter 6.

4.1 Metrics

4.1.1 ROUGE

In our research, we used the ROUGE metric, which compares the quality of the human and automatic summary (Lin, 2004). We have chosen this metric because it shows good statistical results compared to human judgment on the DUC datasets (Paul Over, 2001; Paul Over, 2002; Paul Over, 2003). Also, part of the previous works we were inspired used this metric (Tan, Wan, and Xiao, 2017b; Tan, Wan, and Xiao, 2017a; Cohan et al., 2018; Nallapati, Zhai, and Zhou, 2016; Gavrilov, Kalaidin, and Malykh, 2019). In ROUGE, a reference summary is a summary written by people, while the hypothesis (candidates) summary is an automatic summary. When calculating ROUGE, we can use several references summary; this feature makes more versatile comparisons (than comparing with only one reference summary). This metric bases on algorithms that use n-gram overlapping: the ratio of the number of overlap n-gram (between reference and candidate) to the total number of the n-gram in the reference. ROUGE is a toolkit of metrics, and we used ROUGE-N and ROUGE-L in the current research. We will introduce these metrics in more detail below.

$$\text{ROUGE-N} = \frac{\sum_{S \in \{\text{ReferemceSummaries}\}} \sum_{\text{gram}_n \in S} \text{Count}_{\text{match}}\left(\text{gram}_n\right)}{\sum_{S \in \{\text{ReferemceSummaries}\}} \sum_{\text{gram}_n \in S} \text{Count}\left(\text{gram}_n\right)} \tag{4.1}$$

Where n stands for the length of the n-gram, $gram_n$, and $Count_{match}(gram_n)$ is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries (Lin, 2004). We used a particular case of ROUGE-N: ROUGE-1, and ROUGE-2.

$$\begin{aligned} R_{lcs} &= \frac{LCS(X,Y)}{m} \\ \text{ROUGE-L: } P_{lcs} &= \frac{LCS(X,Y)}{n} \\ F_{lcs} &= \frac{\left(1+\beta^2\right) R_{lcs} P_{lcs}}{R_{lcs} + \beta^2 P_{lcs}} \end{aligned} \tag{4.2}$$

Where $LCS(X,Y)$ is the length of a longest common subsequence of X and Y and $\beta = \frac{P_{lcs}}{R_{lcs}}$ (Lin, 2004)

ROUGE metric has several limitations:

- This metric is based on the overlapping of general content, so it does not take into account grammaticality and fluents;
- An abstractive summary can be expressed in other words, sentences, and general material while keeping the main ideas as in a reference summary. In this case, the ROUGE metric will not be able to detect it and will show a low summarization result;

4.1.2 Cross-entropy

The basic concept in Cross entropy is Shannon information (Shannon, 1948), which measures the number of bits by which a message can be encoded. Shannon's information measures the amount of surprise information in a message: if the message is rare, then it has more surprises than the message often. A quantitative value of the amount of information requires the use of probabilities; therefore, the relation of information theory to probability.

$$I = -log_2(P(x)) \tag{4.3}$$

The information measure can be called a bit (log with a base of 2) or "nuts" (log with a base of e or natural logarithm).
Entropy (Shannon entropy) (MacKay, 2003) - this metric measures the average size of a message/event getting from a random variable. It answers the question: how many bits do we need (in the case of log_2) to minimize the encoding of messages from the distribution of X. If the distribution of the random variable X is skewed, then it will have a low entropy, while in a balanced distribution the entropy will be high (as in this distribution of surprise more).

$$H(X) = \mathbb{E}[I(x)] = -\mathbb{E}[\log P(x)] \tag{4.4}$$

Cross entropy is a metric that measures the difference between two distributions of random variables. It acts as a loss function in classification models like logistic regression and neural networks.

$$H(P, Q) = -\sum_{i=1}^{N} P\left(x_i\right) \log Q\left(x_i\right) \tag{4.5}$$

In classification problems, we encode each class as a vector (one-hot encoding). This vector P:

- Has a size equal to the number of classes;
- Codes in each position a class;
- Contains zeros in all positions and one in a position with a class;

This vector has an entropy of zero. Next, we compare two vectors: P and Q (the one that generated the model) using the cross-entropy function. The closer the metric tends to zero, the more our classes are similar. Thus, the goal is to minimize cross-entropy.

4.2 Extractive approach

The extractive approach is one of the techniques for solving summarization tasks. It selects words and sentences from the source text that best describe the entire document or several documents; The algorithm combines the selected entities to create the resulting summary.
Most texts contain the main idea in the first sentence, called the topic sentence. The topic sentence contains essential information from a document (so-called 5W1H information, where 5W1H stands for who, what, where, when, why, how)(Gavrilov, Kalaidin, and Malykh, 2019). Models developed on this assumption show good results Gavrilov, Kalaidin, and Malykh, 2019. However, Tan, Wan, and Xiao, 2017c affirms that the main idea of text distributes across different sentences of the entire document.
In this work, we used graph-based algorithms for extractive summarization: TextRank (Mihalcea and Tarau, 2004) and LexRank (Erkan and Radev, 2011) bases on the PageRank algorithm (Page et al., 1999).

4.2.1 PageRank

PageRank (Page et al., 1999) is an algorithm developed by Google [1] and used to ranking search results. The idea of the algorithm is that for a specific page, the algorithm calculates the PageRank value based on pages that link to the current one. PageRank is a graph algorithm: nodes are web pages, edges are links between them; each edge has a weight that takes into account the number of links on the page and the weight of all incoming links.

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)} \tag{4.6}$$

where $PR(u)$ - value of PageRank of page u, B_u - all the pages that link to page u, $PR(v)$ - value of PageRank of page v and $L(v)$ - the total number of links from page v.

4.2.2 TextRank

TextRank - is a graph-based ranking model for text processing, developed by Rada Mihalcea and Paul Tarau (Mihalcea and Tarau, 2004). In automatic summarization problems, TextRank applies as an algorithm that extracts more "representative" sentences for the given text. This algorithm is similar to PageRank and has the following differences:

- Instead of web pages, it uses sentences or phrases inside the document;
- It uses an only undirected graph;
- Instead of probabilistic similarity, it uses metrics: BM25, TF-IDF;

BM25 is a bag-of-words retrieval algorithm, proposed (Robertson and Zaragoza, 2009), that ranks a set of documents based on the query terms appearing in each document, regardless of their proximity within the document.

[1] https://about.google/

Given a query Q, containing keywords $q_1, ..., q_n$, the BM25 score of a document D is:

$$\text{score}(D, Q) = \sum_{i=1}^{n} \text{IDF}\left(q_i\right) \cdot \frac{f\left(q_i, D\right) \cdot \left(k_1 + 1\right)}{f\left(q_i, D\right) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)} \tag{4.7}$$

where $f(q_i, D)$ is term frequency for q_i in the document D, $|D|$ is the number of words in the document D, and *avgdl* is the average document length in the entire text collection. IDF(q_i) is the inverse document frequency weight of the query term q_I, computed as:

$$\text{IDF}\left(q_i\right) = \log \frac{N - n\left(q_i\right) + 0.5}{n\left(q_i\right) + 0.5} \tag{4.8}$$

where N is the total number of documents, and $n(q_i)$ is the number of documents containing q_i.

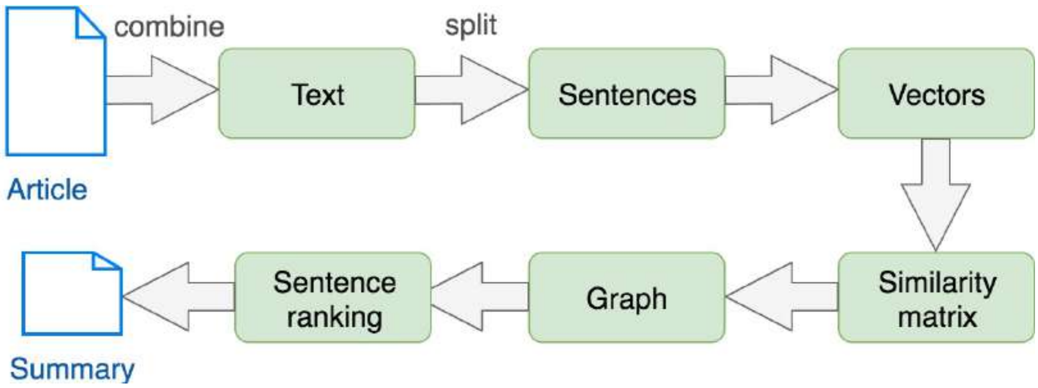

FIGURE 4.1: Stages of TextRank algorithm.

The figure 4.1 shows all the stages of the algorithm. The resulting summary is obtained through a combination of top-ranking sentences or length cutoff to limit the size of the summary. Measurement for the similarity between sentences:

$$\text{Similarity } \left(S_i, S_j\right) = \frac{\left|\left\{w_k | w_k \in S_i \& w_k \in S_j\right\}\right|}{\log\left(\left|S_i\right|\right) + \log\left(\left|S_j\right|\right)} \tag{4.9}$$

4.2.3 LexRank

Lexrank (Erkan and Radev, 2011) is another graph-based algorithm applied to text summarization problems. This algorithm also selects the most important sentences like the previous one, but has some differences from TextRank:

- uses cosine TF-IDF as a measurement for the similarity between sentences;
- Uses unweighted edges applying after the threshold, unlike TextRank (uses similar scores as weights);
- Can create a summary from several documents (TextRank from only one document);

Measurement for the similarity between sentences:

$$\text{idf-modified-cosine}\left(x, y\right) = \frac{\sum_{w \in x, y} \text{tf}_{w,x} \text{tf}_{w,y} \left(\text{idf}_w\right)^2}{\sqrt{\sum_{x_i \in x} \left(\text{tf}_{x_i,x} \text{idf}_{x_i}\right)^2} \times \sqrt{\sum_{y_i \in y} \left(\text{tf}_{y_i,y} \text{idf}_{y_i}\right)^2}}$$

4.3 Word embedding

To solve some of the problems in NLP needs to represent the words in the form of a vector; such a representation can be used to obtain features and mathematical operations with words (vector representation of words). There are several approaches to how to present a word as a vector. The simplest is called one-hot encoded (we described its properties in the Cross entropy section). The disadvantages of this approach are:

- Vector contains a large number of zeros
- The vector has a huge dimension (equal to the size of the dictionary)

These shortcomings increase the complexity of working with such vectors. Words in the form of a vector can be represented more compactly (Harris, 1954) (mostly from 100 to 300 dimensions).
In this section, we will introduce several techniques that help us get the word embedding in a more concise format: Word2vec (Mikolov et al., 2013a; Mikolov et al., 2013b) and FastText (Mikolov et al., 2017). These techniques solve several problems: embedding a vector in the compact form with real numbers and support semantic, syntactic similarity, and relation with other words.

4.3.1 Word2vec

Word2vec is a group of models that are used to embedding words. Word2vec consists of shallow two-layer neural networks. They are trained to restore the linguistic context of the dictionary. Word vectors are placed vector space; each unique word in the corpus is assigned a corresponding vector in space. Thus, they all have a shared context in the body, are located close to each other in space (Mikolov et al., 2013b), an example of which is shown on 4.2.
Word2vec uses two ways to get embeds: Skip-Gram and Common Bag Of Words (CBOW) shown on 4.3. Bag Of Words (CBOW) Mikolov et al., 2013b is a model that receives the context of a word and tries to predict the target word. We send part of the text (5-10 words) to the model's input, except for the center word to be predicted. Words are encoded in a one-hot encoded way.
Skip-Gram Mikolov et al., 2017 gets the word and tries to predict the context (probability distributions) (turned backward than Bag Of Words (CBOW)). We send the word to the input of the model; at the output, we obtain the probability distribution of the words that are the context of the word. Skip Gram works well with a small data set and finds a good representation of rare words. Bag Of Words (CBOW) is quick and suitable for frequent words.

4.3.2 FastText

FastText allows training supervised and unsupervised representations of words and sentences(Subedi, 2018). FastText supports Skip Gram and Common Bag Of Words (CBOW) training as Word2vec. The application of the pre-trained models to vectorize words has a significant drawback: in these models, rare words may be missing, which leads to an out-of-vocabulary (OOV) problem. Models can be trained in the one domain area, and used in another. FastText helps solve the OOV problem.
FastText uses a skip-gram model training at the n-gram level(Bojanowski et al., 2016). Each word is represented as the sum of the symbolic representations of the

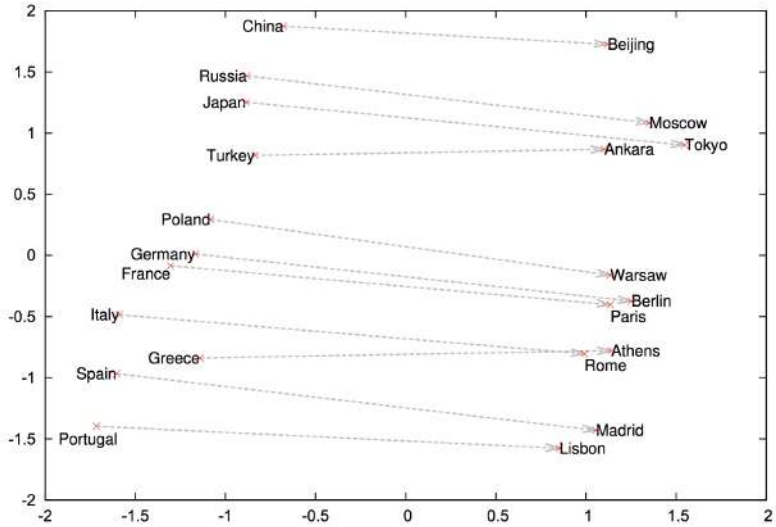

FIGURE 4.2: Two-dimensional PCA projection of countries and their capital cities. The figure illustrates the ability of the model to learn implicitly the relationships between them without any supervised information about what a capital city means. Source: (Mikolov et al., 2013a)

n-gram components of this word. Such an approach makes it easy to represent any OOV word.

4.4 Abstractive approach

An abstractive approach, another type of summarization, create new sentences from which form a summary. Unlike the previous extractive approach, it generates more fluent and grammatically correct sentences, which does not depend on the size of the source text. Abstractive summation methods can be divided into the structure- and semantic-based ones. The idea of structure-based techniques is using prior knowledge, such as trees, ontologies, templates, extraction rules, to encode the data. The semantic-based approaches use the linguistics view of the document to feed into a natural language generation system, with the main focus lies in identifying the noun and verb phrases (Helen, 2018).

4.4.1 RNN

A recurrent neural network (RNN) is a type of artificial neural network that has a cyclic structure and memory. The cyclic structure means that each step has the same parameters, takes a value from the source data and information from the previous step. Thus, the network transfers information from the previous steps, and this is the so-called "memory" of the network. RNN has a significant advantage over feedforward neural networks; the network can process data of any length. When creating feedforward neural networks, the size of the input and output data are constants, they cannot be dynamically changed, so the network requires data of a specific size. RNN, due to its cyclic structure, can process data of different lengths, so it is well suited for processing sequential data. Figure 4.4 illustrates an RNN.

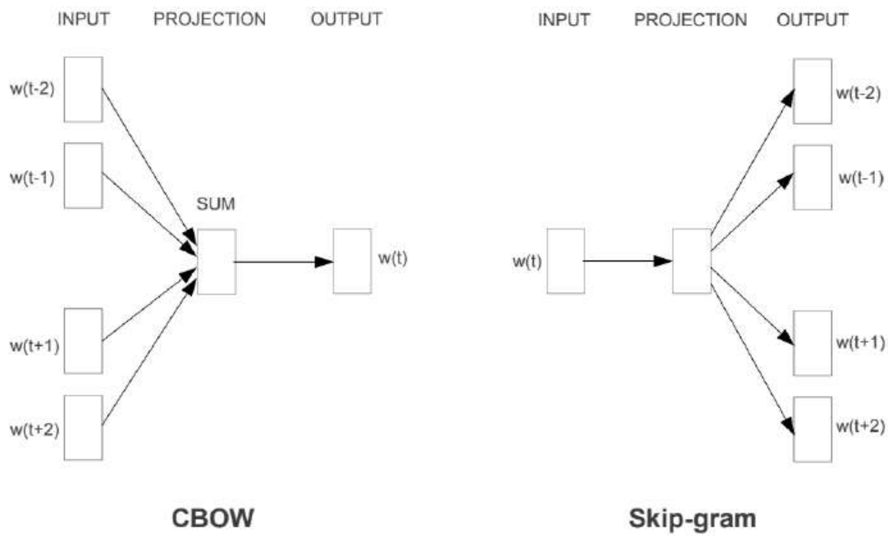

FIGURE 4.3: The CBOW architecture predicts the current word based on the context, and the Skip-gram predicts surrounding words given the current word. Source: (Mikolov et al., 2013b)

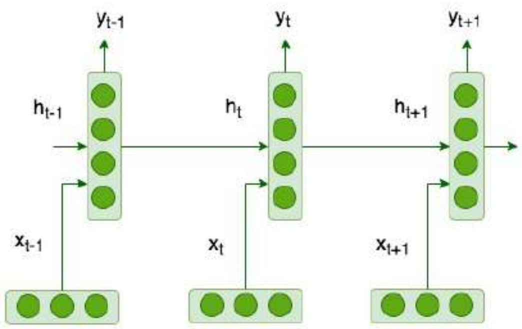

FIGURE 4.4: A Recurrent Neural Network(RNN). Three time-steps are shown.

$$
\begin{aligned}
h_t &= \sigma\left(W^{(hh)} h_{t-1} + W^{(hx)} x_{[t]}\right) \\
\hat{y}_t &= \operatorname{softmax}\left(W^{(S)} h_t\right)
\end{aligned}
\tag{4.10}
$$

$W(hh)$, $W(hx)$ - network weights that are repeated at each step, so the size of the parameters do not depend on the length of the input sequence. The loss function used in RNNs is cross entropy shown in 4.10

$$
J = -\frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{|V|} y_{t,j} \times \log\left(\hat{y}_{t,j}\right) \tag{4.11}
$$

Perplexity 4.12 is a measure of confusion where lower values imply more confidence in predicting the next word in the sequence (compared to the ground truth outcome).

$$
\text{Perplexity } = 2^J \tag{4.12}
$$

RNN has its advantages and disadvantages. The pros: the network can process sequences of different lengths, the size of the model does not increase with increasing incoming sequence, and the model can use the information from many backward steps (in theory). The cons: it is impossible to train the model in parallel, and in practice, it is difficult to obtain information from many steps backward.

4.4.2 LSTM

Long short-term memory (LSTM) is an extended RNN (Hochreiter and Schmidhuber, 1997). It intends to solve the significant problem with RNN: vanishing or exploding gradient. During training, the model uses the backpropagation algorithm to propagate an error back through the cyclical connection. RNN models use the same weight matrix, and the number of layers (equal to the size of the input sequence) can be enormous - this all leads to the problem described by (Hochreiter et al., 2001). Backpropagates error signals strong depend on the weights. If the weight is greater than 1, then explode occurs, if less than the vanishing gradient. LSTM was designed to have a more persistent memory. Thus, the model can save term patterns in the inner layers for the long or short time and does not suffer from exploding or vanish gradient.
At the time step t model uses a previous value of a hidden state $h(t1)$, a memory cell $c(t1)$ and an input vector x(t). The new value of memory cell uses gates to forget a part of the previous value and remember a new value. Full equation showed in 4.134.5 and showes detailed internals of a LSTM.

$$
\begin{aligned}
i_t &= \sigma\left(W^{(i)}x_t + U^{(i)}h_{t-1}\right) && \text{(Input gate)}\\
f_t &= \sigma\left(W^{(f)}x_t + U^{(f)}h_{t-1}\right) && \text{(Forget gate)}\\
o_t &= \sigma\left(W^{(o)}x_t + U^{(o)}h_{t-1}\right) && \text{(Forgetire gate)}\\
\tilde{c}_t &= \tanh\left(W^{(c)}x_t + U^{(c)}h_{t-1}\right) && \text{(New memory cell)}\\
c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t && \text{(Final memory cell)}\\
h_t &= o_t \circ \tanh\left(c_t\right)
\end{aligned}
\tag{4.13}
$$

4.4.3 Sequence-to-sequence model

Sequence-to-sequence (seq2seq) (Sutskever, Vinyals, and Le, 2014) is a language model that learns the probability distribution between an input and an output sequence. That is, during training, seq2seq maximizes the log-likelihood $P(y|x)$, where x is the input sequence, and y is the output. The model consists of two RNN models: an encoder and a decoder. The encoder receives a sequence (for example, tokens in a sentence) and encodes it to a vector that represents an input sequence ("context vector"). The decoder receives a "context vector" from the encoder, sets it as an initial hidden state, and generates an output sequence; the decoder generates a sequence until it faces an end sentence token. As an encoder, we can apply RNN based models, likes RNN or LSTM, as well as biRNN or biLSTM.
In the recent past, sequence-to-sequence models have been successful in such problems as machine translation (Bahdanau, Cho, and Bengio, 2014), speech recognition (Chorowski et al., 2015) and video captioning (Venugopalan et al., 2015)

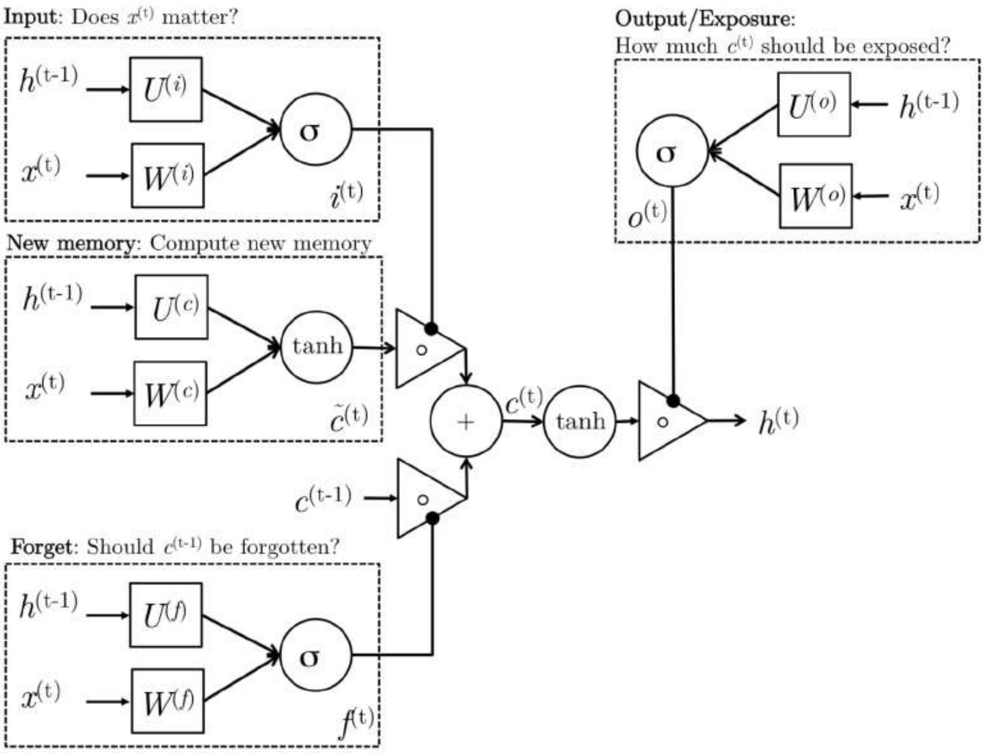

FIGURE 4.5: The detailed internals of a LSTM. Source: (*CS224n: NLP with Deep Learning*).

4.4.4 Neural Machine Translation

Neural machine translation is one of the machine translation approaches proposed (Kalchbrenner and Blunsom, 2013; Sutskever, Vinyals, and Le, 2014) and improved upon using attention mechanism (Bahdanau, Cho, and Bengio, 2014; Luong, Pham, and Manning, 2015).
From a probabilistic point of view, the task of Neural machine translation is to find model parameters such that maximize the conditional probability of sentence pairs using a parallel training corpus. Most of the proposed model consists of an encoder and decoder, which represent the source and output languages. The encoder model reads the input data, translates it into a fixed vector. Decoder, receiving a vector from an encoder, generates output. Unlike statistical machine translation (SMT), encoder and decoder in this model train together to maximize the likelihood of a correct translation. NMT has now become a widely applied technique for machine translation, as well as a practical approach for other related NLP tasks such as dialogue, parsing, and summarization.

4.4.5 Attention mechanism

Our brain almost always selects only essential information to solve a problem. That is the main idea of the Attention mechanism. The sequences of tokens that enter the input of the seq2seq model have different significance. However, the seq2seq model encodes in the final "context vector" all tokens from the incoming sequence as if they were equal in significance for the entire sequence. (Bahdanau, Cho, and Bengio, 2014) saw a weakness of this approach and declared that different parts of the incoming sequence have different levels of significance or importance.

Specifically, in the encoder-decoder model, the Attention mechanism allows to solve this problem as follows: it allows the decoder to select and use the most significant tokens at a certain point in time from the incoming sequence. The attention mechanism is shown on 4.7 and equations 4.14-4.18.

$$p\left(y_i \mid y_1, \ldots, y_{i-1}, \mathbf{x}\right) = g\left(y_{i-1}, s_i, c_i\right) \tag{4.14}$$

$$s_i = f\left(s_{i-1}, y_{i-1}, c_i\right) \tag{4.15}$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j \tag{4.16}$$

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k=1}^{T_x} \exp\left(e_{ik}\right)} \tag{4.17}$$

$$e_{ij} = a\left(s_{i-1}, h_j\right) \tag{4.18}$$

where s_i is an decoder hidden state, h_j vectos of annotation (produced from encoder), c_i is context vector (with attention mechanism), a is an alignment model, which parameterized as a feedforward neural network and is jointly trained with all the other components of the proposed system (Bahdanau, Cho, and Bengio, 2014).

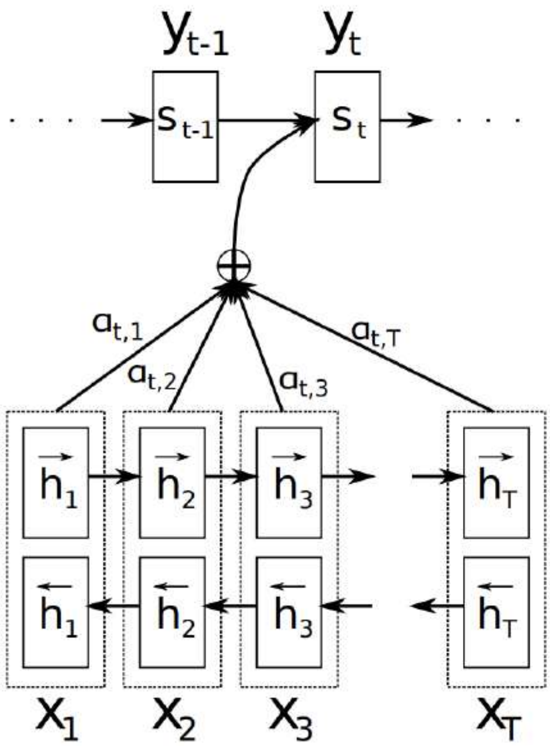

FIGURE 4.6: The graphical illustration of the proposed model trying to generate the t-th target word yt given a source sentence $(x_1, x_2, ..., x_T)$. Source: (Bahdanau, Cho, and Bengio, 2014).

4.4.6 Transformers

Despite success in solving problems of RNN/LSTM based models, they have certain disadvantages that prevent them from developing further. Firstly, these models are not hardware friendly and require many resources, while training compared to

other models (Culurciello, 2019). The second drawback of models is their sequential nature, prevent parallelization, which becomes critical at longer sequence lengths, as memory constraints limit batching across examples (Vaswani et al., 2017). The third drawback is that in the RNN based model, broadcasting signals at long distances disappear (self-attention link).
Self-attention mechanism and convolution neural approaches help to solve these problems; they form transformers. A transformer is a neural networks model architecture that excludes the RNN based approach and consists only of an attention mechanism to catch global dependencies between input and output (Vaswani et al., 2017). These models consist of an encoder and decoder, which include stacked self-attention, point-wise, and fully connected layers. The figure 4.7 shows the architecture of the transformer proposed (Vaswani et al., 2017).
These architectures allow training models in parallel, which increases the quality and accuracy of models and also shows good results in language understanding problems (Uszkoreit, 2017)

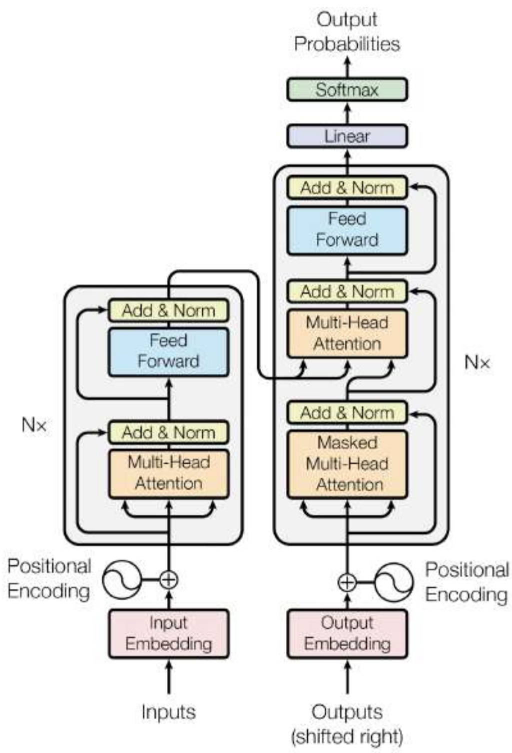

FIGURE 4.7: The Transformer - model architecture. Source: (Vaswani et al., 2017).

4.4.7 BERT

BERT (Bidirectional Encoder Representations from Transformers) is a language representation model that has the architecture of a transformer model (Vaswani et al., 2017). The main feature is that BERT can present the text as deep bidirectional representations, in contrast to other language representation models (Peters et al., 2018; Radford, 2018).
During train, the model uses two tasks: predict the original vocabulary id from random masked word and predict the next sentence (model receives two sentences and tries to predict if the second sentence from the original document). Another feature

of BERT is that it is quite easy to use as a pre-trained model for solving downstream NLP tasks; add a fine-tuning layer and train the model on the data from the downstream task. The figure 4.8 shows the pre-training and fine-tuning procedures for BERT model. The using BERT as a pre-trained model creates state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications (Devlin et al., 2018).

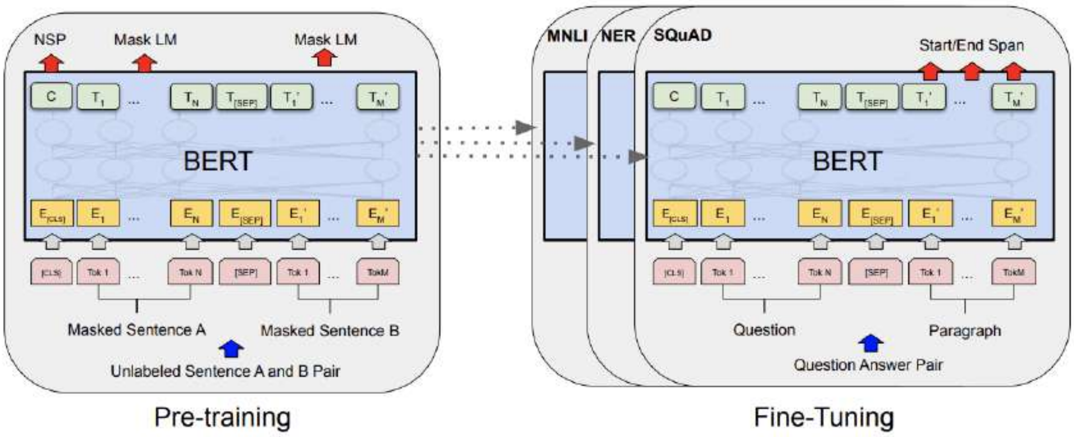

FIGURE 4.8: The figure shows pre-training and fine-tuning procedures for the BERT model. The same architectures and model parameters (to initialize models for different down-stream tasks) are used in both pre-training and fine-tuning stages. During fine-tuning, all parameters are fine-tuned. Source: (Devlin et al., 2018).

4.5 Differences in broadcasts styles and news

In the current chapter, we will discuss one of the features of our data corpus: broadcasts and news belong to different styles of language. Broadcasts in our dataset are a monologue of the speaker that describes the sports competition. This monologue refers to the spoken style and shares the following properties: emotionality, using incomplete or short sentences, introductory words, interjections, modal particles, frequent repetitions to emphasize ideas, and others. Even though the author describes events during the game, he also shares his thoughts and feelings in the comments. Non-linguistic factors play an essential role in speech: facial expressions, gestures, the environment; however, we cannot detect this using the text version of the broadcast.

News, in turn, refers to a different style, called publicists, whose main task is to communicate information about some event, influence the masses of people, and form a specific viewpoint to public events. News stories also contain at least one of the following important characteristics relative to the intended audience: proximity, prominence, timeliness, human interest, oddity, or consequence[2]. We also want to note that some of the news from our dataset are participants' comments regarding the game. These comments cover the analysis of the game and the results; most of the comments are quite emotional and cover the general situation on the one hand only.

We have presented these differences in order to emphasize that the ROUGE metric and extractive approaches can give irrelevant results. Extractive approaches will

[2] https://en.wikipedia.org/wiki/News_style

form the news as part of the broadcast, that is, news written in a conversational style. While the "gold" news belongs to the news style and consists of different n-grams (the unit of measure for ROUGE) than the broadcast.

Chapter 5

Model

In the current chapter, we briefly introduce the approaches and frameworks that were used in our experiments. These approaches are based on Neural machine translation (NMT). First, we discuss the OpenNMT framework (Klein et al., 2018) and later about the approaches proposed (Liu and Lapata, 2019), which we called PreSumm. The fundamental components of these approaches, we described in the previous chapter 4, in the next chapter 6, we will describe the experiments and results using these and other approaches.

5.1 OpenNMT

In our research, one of the frameworks that we use is OpenNMT. OpenNMT is an open-source framework that supports NMT research. This framework has a clear modular architecture and functional extensibility. It is implemented in the LUA and python programming language; it imported to PyTorch (we used this module for experiments)[1] and TensorFlow [2].
We use OpenNMT for NMT experiments because it includes many parameters and features for tuning our models:

- gated RNN such as an LSTM (Hochreiter and Schmidhuber, 1997) or GRU (Chung et al., 2014) which help the model learn long-term features;
- stacked RNNs, which consist of several vertical layers of RNNs (Sutskever, Vinyals, and Le, 2014);
- input feeding, where the previous attention vector is fed back into the input as well as the predicted word, is quite helpful for machine translation (Luong, Pham, and Manning, 2015);
- beam search, which considers multiple hypothesis target predictions at each time step. (Klein et al., 2017b)

Chapter 6.4 describes how we apply OpenNMT to solve our problem.

5.2 PreSumm

Another approach that we utilize in our research proposed (Liu and Lapata, 2019) called PreSumm. This approach shows better results compared to OpenNMT.
Yang Liu and Mirella Lapatas in (Liu and Lapata, 2019) propose to encode the whole document, keeping its sense, to generate a compact conclusion. The authors scale

[1]https://github.com/OpenNMT/OpenNMT-py
[2]https://github.com/OpenNMT/OpenNMT-tf

down the problem of extractive summarization to the problem of categorization: needs to take sentences that best show the main thoughts/ideas of the whole text. This approach is a supervisor learning: the model should know which sentences most contain the primary sense. The authors solved this problem using the greedy approach model (Nallapati, Zhai, and Zhou, 2016) - choose sentences from the source text that maximize metric ROUGE (Lin, 2004); the algorithm stops when adding sentences does not increase ROUGE. During the generation of summary, the model sets a rating to the selected offers and selects the top 3 sentences; the original summary forms these sentences.
The abstractive summarization is reduced to the NMT approach: an encoder contains a trained model (BERT), and a decoder contains a randomly initialized BERT. If training two models - one pre-trained and other randomly initialized - at the same time with the same parameters, then one model can be overfitting, and the second can underfitting, or vice versa. The authors use different training parameters (learning rates and warmups). BERT was trained on two tasks: masked generating token (ID from vocab) and "next sentence pair". Output vectors are based on tokens.
The sequence provided to the BERT input (source) should contain the [CLS] token (into which BERT encoded the results for categorical tasks), and contain [SEP] token divides the sentences inside the sequence. Source tokens also contained information about their position in the entire sequence and information to which sentence (part of the sequence) the token belongs.
The authors made some changes around the BERT model and called it BERTSUM. They (1) added the [CLS] token at the beginning of each sentence and (2) changed the segmentation embeddings to separate several sentences in one sequence; if the document consists of $[sent1, sent2, sent3, sent4, sent5]$ then the following segmentation embeddings $[E_A, E_B, E_A, E_B, E_A]$ will be assigned to the tokens inside the sentences. Figure 5.1 shows the BERT and BERTSUM models.

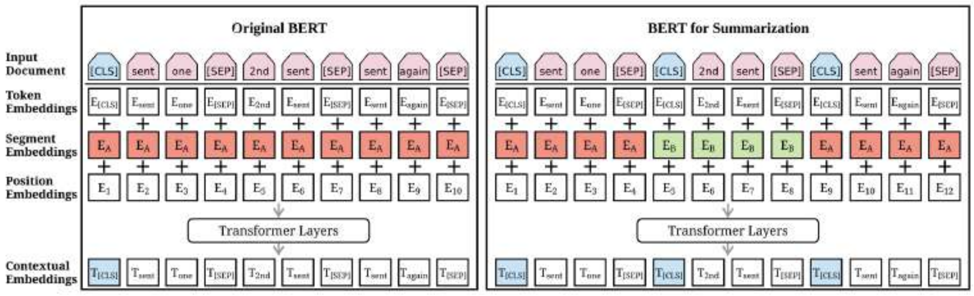

FIGURE 5.1: Architecture of the original BERT model (left) and BERTSUM (right). Source: Liu and Lapata, 2019

BERTSUMEXT is a model for extractive summarization. It is inherited from BERTSUM, uses two layers of transformers, and its final layer is a sigmoid classifier (for solving problems of classifying sentences). BERTSUMABS is a model for abstractive summarization. This model uses the NMT approach, pre-trained BERTSUM as an encoder, and the randomly initialized transformer in the decoder. BERTSUMEXTABS is also using the NMT approach; only unlike BERTSUMABS, it uses the pre-trained BERTSUMEXT on the extractive summarization task as an encoder. This approach has an implementation [3], but there are some limitations to it:

[3]https://github.com/nlpyang/PreSumm

- The code was written to work only with datasets from paper (CNN/DM, NYT, XSumm) (**NYT**; Hermann et al., 2015b; Sandhaus, 2008; Narayan, Cohen, and Lapata, 2018);
- Code works with English sequences only;
- Uses only the fixed version of BERT (bert-base-uncased)

Chapter 6.5 describes how we got rid of these restrictions.

Chapter 6

Experiments

In this chapter, we conduct a series of experiments with different approaches and evaluate their quality. We start by applying extractive methods to solve our problem, using different implementations of PageRank algorithms, as well as TextRank and LextRank approaches; after we experiment with the Oracle model. Next, we will use abstractive methods, in particular, the NMT approach. First, we will use the OpenNMT framework with different types of encoders and other parameters. Later we switch to PreSumm approaches and experiment with different parameters, types of pre-trained models. We will also implement a series of experiments with the augmentation of our text data. Finally, we conduct a human evaluation to judge the quality of our results.

6.1 TextRank approaches

Implementation details

In this series of experiments, we will apply two models to the implementation of the TextRank algorithm, the first based on the PageRank [1] algorithm, the other on the Gensim Textrank[2]. These approaches differ in the similarity function of two sentences. The PageRank-based model uses cosine distance between vectors of sentences (below we describe the algorithm of getting vector of the sentence); the Gensim model based on the BM25 algorithm, which we described in chapter 4.2.2.
For the PageRank model, we preprocessed the broadcast text. We split the text into sentences using NLTK(Loper and Bird, 2002), then into words, and lemmatized each word using pymystem3 (Segalovich, 2003). To vectorize the words, we used two different pre-trained models: the word2vec model, trained on the Russian National Corpus (Kuratov and Arkhipov, 2019), and the FastText model, trained on the news corpus (Shavrina T., 2017). To vectorize one sentence, we added all word vectors in a sentence together and divided by the count of words. Then, we calculated the cosine distance between all sentences, build a similarity matrix, converted it to a graph, and applied the PageRank algorithm. The Pagerank parameters remained by default from the library. After that, we selected sentences with a maximum page ranking and formed a summary from them.
For the Textrank model, we used raw broadcast text and parameter $ratio = 0.2$, which adjust the percentage size of the summary compared to the source text (20% of the sentences from the source text will be in summary).

[1] http://bit.ly/diploma_pagerank
[2] https://radimrehurek.com/gensim/index.html

Results and examples

The TextRank approaches results are represented in 6.1 and samples in A.3.

method \ROUGE	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
PageRank W2V	0.06	0.15	0.08	0.	0.	0.	0.06	0.15	0.06
PageRank FT	0.06	0.13	0.08	0.	0.	0.	0.06	0.13	0.06
TextRank Gns	0.05	0.17	0.08	0.	0.01	0.	0.05	0.16	0.06

TABLE 6.1: ROUGE scores of TextRank approaches.PageRank W2V and PageRank FT models based on PageRank and used word2vec and FastText models. TextRank Gns - an algorithm from the Gensim toolkit.

All algorithms from the TextRank experiments showed very similar results: ROUGE-1 F1 is the same in all algorithms. ROUGE-2 showed the worst results among all ROUGE metrics. TextRank works better than PageRank(W2V) and PageRank(FT): ROUGE-1 F1 more than 0.01. However, the results remain not significant: the TextRank algorithm has the maximum ROUGE F1, which is 0.08.
We assume that we got such results because different people describe sports commentary and news in different formats, styles and situations: the commentator describes the emotionally sporting game online, with details; the author of the news, calmly and dryly reports the results or takes an interview from the game player or coach. Therefore, these texts, when comparing, use different words, word forms, expressions. This property leads to an insignificant ROUGE metric based on the overlapping of common words.
We also compared the generated summary (news) with two types of "gold" news: before and after the sports game. The ROUGE metric showed an average of 0.03 more at the news after the match than before the match. We assume the next: news, published after the sports game, more describe the games and the results (similar to the broadcast), than the news published before the sport game.

6.2 LexRank

Implementation details

In the next experiment, we used another extractive approach called the LexRank algorithm. We described the difference between LexRank and TextRank in 4.2.3. We found a flaw in the library: the IDF algorithm workes for a very long time with extensive data (from our experiments). We re-write that part of the algorithm [3]. We chose the top 10 offers, the rest of the parameters used by default[4].

[3] https://github.com/DenisOgr/lexrank/pull/1/files)
[4] https://pypi.org/project/lexrank/

Results and examples

method \ *ROUGE*	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
LexRank	0.07	0.1	0.08	0.	0.	0.	0.07	0.09	0.06

TABLE 6.2: ROUGE-1/ROUGE-2/ROUGE-L score of LexRank results.

Петров обыграл сразу двух игроков на левом фланге атаки "Астон Виллы", отдал Карью, но тот замешкался на углу вратарской. Хозяева поля заработали штрафной ближе к левому флангу атаки, Эшли Янг пробил неплохо, но Акинфеев вытащил мяч в нижнем углу ворот! Алдонин пробил издали довольно сильно, голкипер "Астон Виллы" не без труда справился с этим ударом, зафиксировав мяч не с первого раза. Petrov beat two players at once on the left flank of the attack, "Aston Villa", gave Carew, but he hesitated at the corner of the goalkeeper. The home team earned a free-kick close to the left flank of the attack, Ashley Young shot well, but Akinfeev pulled the ball in the bottom corner of the goal! Aldonin struck from afar rather strongly, the goalkeeper of "Aston Villa" not without difficulty coped with this blow, fixing the ball not the first time.
ВЕЛЛИТОН! Почти один на один! А "почти" потому, что Дикань своевременно сыграл на выходе. ВЕЛЛИТОН головой (!) наносит удар в упор. Даже Дикань бы не спас.Но мяч прошел рядом со штангой. АлЕкс вполне прилично закручивает мяч со стандарта в центр штрафной. Welliton! Almost one on one! And "almost" because Dikan played timely on the way out. WELLITON's head (!) Strikes point-blank. Even Dikan would not have saved. But the ball went near the post. Alex quite decently spins the ball from the standard into the center of the box.
Муджири исполнял остроумный пас на Сычева. Березовский очень надежен. Ему пока даже защитники не нужны. МОМЕНТ У СЫЧЕВА! Да... Дмитрий, выходя один-на-один, продемонстрировал что губить возможности он может не хуже Муджири. Рахимов едва не выбегает на поле... Davit Mujiri performed a clever pass on Sychev. Berezovsky is very reliable. He does not even need defenders yet. MOMENT AT SYCHEV! Yes ... Dmitry, going one-on-one, demonstrated that he can ruin opportunities no worse than Mujiri. Rakhimov almost runs out onto the field ....

TABLE 6.3: Examples of summaries generated using LexRank approach.

Table 6.2 shows the results of the current experiment, and table 6.3 shows samples of summaries. The experiment with the LexRank approach showed the same result as the TextRank for ROUGE-1 F, more by 0.01 in ROUGE-L F and less by 0.02 in ROUGE-2 F. This algorithm is very similar to TextRank (more details in (4.2.3); therefore their results are pretty close to each other.
In the next experiments (6.3), we will create a model that shows the maximum

ROUGE value that we can get from extractive approaches. We called this model Oracle.

6.3 Oracle

Implementation details

This model generates an extractive summary that has the most value ROUGE between broadcast and news. We used the greedy search algorithm: we found the value of custom rouge (ROUGE-1 f + ROUGE-2 f) between each sentence from the broadcast and all the sentences in the news and selected the top 40 sentences. This algorithm stopped working in one of two cases: (1) the number of sentences is greater than the requested upper threshold (40 sentences) or (2) adding the next sentences does not increase ROUGE. We called this model is Oracle because it generated a summary with the maximum ROUGE that we could get from the extractive approaches. The actual distribution of the number of offers can be seen on the figure 6.1. We can see, that the most number of sentences in the model less than 20.

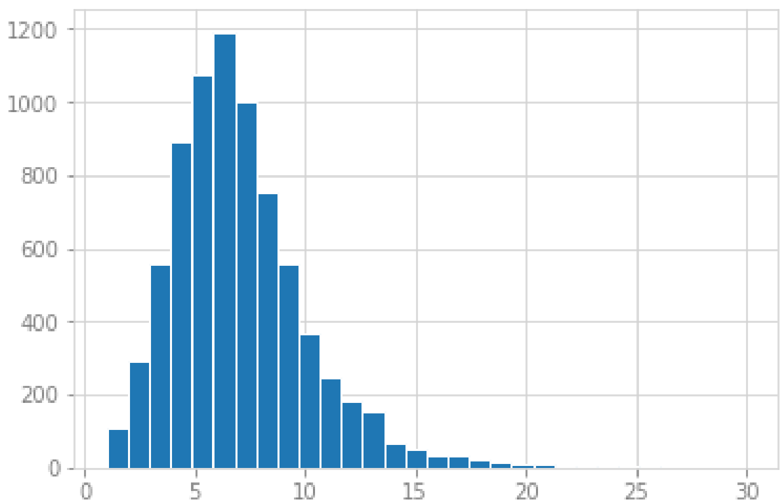

FIGURE 6.1: This figure shows the distributation on number of sentences using Oracle model.

Results and examples

method \ROUGE	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
Oracle	0.2	0.22	0.21	0.02	0.02	0.02	0.18	0.2	0.19
PageRank W2V	0.06	0.15	0.08	0.	0.	0.	0.06	0.15	0.06
PageRank FT	0.06	0.13	0.08	0.	0.	0.	0.06	0.13	0.06
TextRank Gns	0.05	0.17	0.08	0.	0.01	0.	0.05	0.16	0.06
LexRank	0.07	0.1	0.08	0.	0.	0.	0.07	0.09	0.06

TABLE 6.4: ROUGE scores for all exptractive approaches.

Table 6.1 shows the results of the Oracle model. We see a significant increase in ROUGE compared to the previous (6.1, 6.2) approaches: ROUGE-1 F is 0.13 more than other extractive approaches, ROUGE-L f is 0.13, and ROUGE-2 is 0.02 more. We can see that Oracle chose significant comments that would describe the game and its result (see table A.4). We assume that such a low maximum value (of the ROUGE metric) that can be obtained from the extractive model because the authors of the broadcast and news use different words, phrases. We describe this in more detail in the chapter 4.5 In the next experiments (6.4, 6.5), we will use abstractive summation.

6.4 OpenNMT

Implementation details

For the summarization problem in OpenNMT, we used the approach described in the article (Gehrmann, Deng, and Rush, 2018). We used the entire text from broadcast and the shortest news as a source and as a target accordingly. The dataset was divided into three parts: train/test/validate in sizes: 6143/767/760 (90%/10%/10%). We used the following approaches and parameters when training and testing the model [5]:

- the copy-attention mechanism described in the article (Vinyals, Fortunato, and Jaitly, 2015);
- the attention mechanism introduced by Bahdanau, Cho, and Bengio, 2014 instead of that by Luong, Pham, and Manning, 2015;
- shareing the word embeddings between encoder and decoder;
- size of vocabulary equal 50000;
- beam search equal 10 (Neubig, 2017);
- Different types of encoder: RNN/biRNN/LSTM/biLSTM/Transformers (Vaswani et al., 2017);
- bridge layer. It gets finish layer from last encoder hidden state and compute initial state for decoder;
- Dropuot equal 0;
- reusing the standard attention as copy attention;

In the first series of experiments, we used different types of the encoder. The experiments in which the encoder type was: RNN/biRNN/LSTM/biLST did not show significant results. All models did not learn anything and produced results similar to:

> "полузащитник <unk> <unk> в в в матче матче <unk> <unk> <unk> <unk> <unk> <unk>"
> "midfielder <unk> <unk> in the match in the match <unk> <unk> <unk> <unk> <unk>"

The experiment that gave better results (compared to our previous ones based on OpenNMT) is using the transformer in the encoder and decoder. We used the transformer proposed (Vaswani et al., 2017) and the number of layers is 4. We got the following examples:

[5] https://github.com/DenisOgr/news-generations

"чемпионат <unk> "тур примечание: время начала матчей - московское.таблица чемпионата беларусистатистика чемпионата беларуси
"championship <unk> round note: the start time of the matches is Moscow. belarus championship table Belarus statistics"

We did not use the ROUGE metric for these experiments, as the results were visually inappropriate.
In the next series of experiments with OpenNMT frameworks, we decided to try to train a model that could generate news, based on other news. We trained two models in the following ways:

- Different news written for one sport game was submitted to the input and output;
- The same news was sent to the input and output;

These models were trained with parameters from previous experiments and showed poor results; The models did not learn any pattern in data and produced, in the more occurrences, the illogical text or characters.
All experiments an abstractive approach use the OpenNMT framework did not show significant results. Most of the generated texts were not readable and logical, so we did not use the ROUGE metric. We got better results when we used transforms in encoders and decoders; training with transformers was faster. We assume that the reason for the failure of using RNN-based models in long sequences; figure 3.2 shows the distribution of the number tokens in broadcasts. Our further experiments will focus on the following:

- Use transformers instead of RNN based models, as an encoder and decoder
- Decrease incoming sequences, keeping the main events in the broadcast. We will use the extractive-abstractive model in section 6.5.3.

6.5 PreSumm

Implementation details

In this series of experiments, we decided to use the approaches described by (Liu and Lapata, 2019). The authors of the article made implementation (make code on github.com) [6]; however, their implementation partly suitable for our experiments. The implementation has the following limitations that we have fixed [7]:

- Designed only for the using of datasets from experiments (CNN/DailyMail, NYT, XSum). Each dataset was provided in its format, and authors wrote separate handlers and preprocessors for each dataset. Preprocessing is a stage that transforms raw data into a data format for the model. We implemented additional preprocessors that transform from text files (data format of our dataset).
- Authors did not implement sharding data. Sharding is the process of breaking data into smaller datasets. Thus, the model loads small data into memory and uses memory more rationally (economically). Unlike the datasets described in the article, the data from current work was much longer: the average length

[6] https://github.com/nlpyang/PreSumm
[7] https://github.com/DenisOgr/PreSumm

in documents from the (Liu and Lapata, 2019) were 500-600 words, but 1500 tokens from our dataset. We implemented a sharding mechanism, which allowed us to avoid problems with "Memory allocation cannot be possible" in CUDA.

- The solution used only bert-base-uncased as the transformer. We added the ability to use any transformers from the two libraries of transformers: "hugging face/transformers" [8] and "transformers-ru"[9].
- Added saving to TensorBoad [10] to analyze the model training phase.

We cut off long broadcasts and news: the maximum length of the broadcast was 2500, and the length of the news 200. Figure 6.3 shows the distribution lengths of the sequences of tokens for the current experiment for broadcasts and figure 6.2 for news. To train the model, we used the NVIDIA Tesla P100 video card and splited our dataset into shards, with a size of 50 examples. Next, we will describe different experiments with different approaches, parameters. We will show the results and examples of the generated summaries.

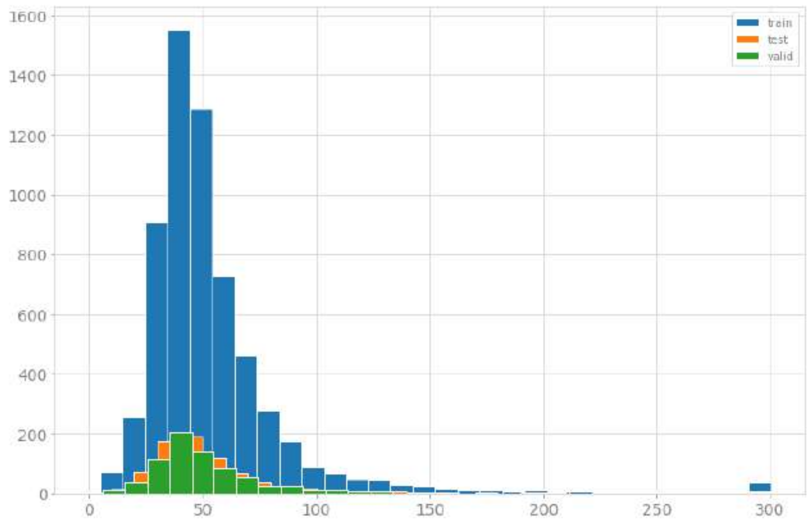

FIGURE 6.2: This figure shows the distributation number of tokens in news (target for Presumm experiments).

[8] https://github.com/huggingface/transformers
[9] https://github.com/vlarine/transformers-ru
[10] https://github.com/tensorflow/tensorboard

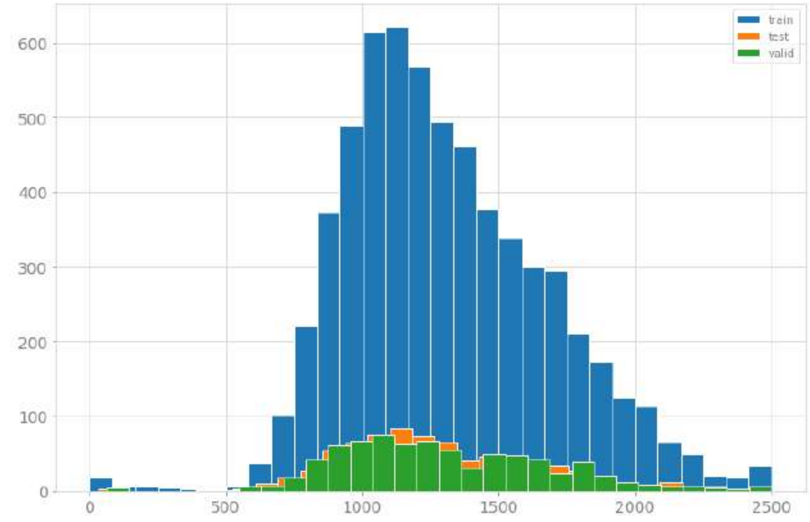

FIGURE 6.3: This figure shows the distributation number of tokens in broadcasts (source for Presumm experiments).

6.5.1 BertSumAbs

In this experiment, we used the abstractive BertSumAbs model with bert-base-multilingual-uncased [11] in its encoder and randomly initialized BERT in the decoder. We trained a model with different learning rates (0.002 for an encoder and 0.2 for a decoder) and warmup (20,000 for an encoder and 10,000 for a decoder) sugested by Liu and Lapata, 2019. The parametr *max_pos* was 512 (default value), and the dropout for decoder was 0.2. We stored the model every 10000 steps and have chosen a model that showed the best results (using ROUGE metric) on the validation dataset.
Results of this experiment are presented in 6.5 and samples in A.1.

method \ *ROUGE*	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
BertSumAbs	0.19	0.25	0.21	0.07	0.1	0.07	0.17	0.23	0.18

TABLE 6.5: ROUGE scores for summaries generated by using model BertSumAbs.

The best ROUGE results show the model that has been trained in 50,000 steps. We noticed that the model tended to overfit after 50 000 iterations. BertSumAbs is the first experiment with Presumm, so we compared the results with previous extractive summation approaches (shown in table 6.4). The ROUGE value in this experiment was the highest compared to all previous experiments: ROUGE-1 F is greater than 0.13, ROUGE-2 F is 0.07, ROUGE-L F is 0.13 more; we made this comparison using the value of the models that showed the highest result, except for the Oracle model (the model is experimental). Then we decided to experiment with another model in the encoder (6.5.2)

[11] https://github.com/google-research/bert/blob/master/multilingual.md

6.5.2 RuBertSumAbs

In this experiment, we used a different model in the encoder, which was trained on Russian texts, RuBERT[12]. The training settings used are the same as in the experiment 6.5.1. The model with the best ROUGE score has been trained in 40,000 steps. Table 6.6 shows the result from experiment RuBertSumAbs and table A.5 show summaries.

method *ROUGE*	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
RuBertSumAbs	0.14	0.26	0.17	0.04	0.1	0.06	0.13	0.24	0.13

TABLE 6.6: ROUGE scores for summaries generated by using RuBertSumAbs.

This model showed lower ROUGE results compared to the BertSumAbs model: ROUGE-1 F is less by 0.04, ROUGE-2 F, and ROUGE-L F are less by 0.01 and 0.05, respectively. We assume that the reason for this is the encoder model: "bert-base-multilingual-uncased" generates contextual vectors better than "RuBERT". In the following experiments, we used only "bert-base-multilingual-uncased" as the base model of the encoder.

6.5.3 BertSumExtAbs

In this experiment, we used the double fine-tune stages for encoder: firstly, we fine-tuning the model to the extractive summarization task, then we fine-tuning that model on the abstractive task (Liu and Lapata, 2019). For the first fine-tune stage, we used BERT "bert-base-multilingual-uncased", learning rate is $2e-3$, dropout is 0.1 *max_pos* is 512, and *warmup_steps* 10000. Next, the trained model was used as an encoder for the abstractive summarization task with the same parameters, as in previous experiments described in 6.5.1.
Results of this experiment are presented in table 6.7 and samples in table A.11.

method *ROUGE*	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
BertSumExtAbs	0.18	0.25	0.2	0.06	0.1	0.07	0.17	0.23	0.17

TABLE 6.7: ROUGE scores for summaries generated by using BertSumExtAbs.

This model has shown results slightly lower than the BertSumAbs, described in 6.5.1. The values of ROUGE-1 F and ROUGE-L F are less than 0.01, and ROUGE-2 F is the same.
Inspecting previous experiments, we realized that *max_pos* - truncates our incoming sequences; the model trains only 512 of the first Broadcast tokens. According to the distribution of token lengths (showed 6.3), this is quite small sequences to getting all vital information from the broadcast. We increased this parameter to 2500, and this led to a problem with the memory allocated problem (we used NVIDIA Tesla P100 16 Gb). In the next experiments, we increased to the maximum (regard for our computation power) *max_pos*=1024 (6.5.4).

[12]http://docs.deeppavlov.ai/en/master/features/pretrained_vectors.htmlbert

6.5.4 BertSumAbs1024

In this experiment, we used training parameters from previous experiments(6.5.1), only increased *max_pos* to 1024 and "bert-base-multilingual-uncased" as an encoder model. The model showed better results (showed in table 6.8), compared with previous models with *max_pos*=512. Model, trained 30 000 steps showed the best results. We made the hypothesis that we need to select sequences of higher dimensions or reduce the size of the input sequences while preserving the essential meanings and ideas of the entire broadcast. Examples of summaries shown in table A.6.

method *ROUGE*	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
BertSumAbs1024	0.2	0.25	0.21	0.07	0.1	0.08	0.18	0.23	0.18

TABLE 6.8: ROUGE scores for summaries generated by using BertSumAbs1024.

The model in this experiment did not show significant improvements, compared to the best models, where we used *max_pos*=512. The values of ROUGE-1 F and ROUGE-2 F are higher by 0.001, and ROUGE-L F is 0.006. We want to note that this model was trained for 30,000 steps, and this is 20,000 steps less than 6.5.1. We have seen that increasing the input sequence from 512 to 1024 did not produce significant improvements, according to the ROUGE metric. We assume that this property of ROUGE metric: the overlap words between summary and "gold" news do not increase while increasing the input sequence in Presumm approaches. In further experiments, except for the Oracle model from section 6.5.5, we will use *max_pos*=1024 to reduce the training time.

6.5.5 OracleA

In this series of experiments, we decided to reduce the incoming sequence (broadcast) by applying the extractive approach techniques. We decided to apply the Oracle model, which selects sentences (in our case, 40 top sentences) from the broadcast, which have the maximum news trip (gold reference). More information about this model described in 6.3. This approach is more experimental: we want to know which ROUGE we could get if we use the best (by the ROUGE standards) extractive summarization technique. In the future, we could replace the extractive model with (6.1) or (6.2).
We used "bert-base-multilingual-uncased" in the encoder. *max_pos* = 512. In this experiment, we trained two models that get a short output (the result of the Oracle model) as an input: (i)OracleA - a model trained with parameters from the 6.5.1 experiment and (ii)OracleEA - model trained with parameters from the 6.5.3 experiment. The best models trained 30000 (BertSumOracletAbs) and 40000 (OracleEA) steps.
Results of this experiment are presented in 6.9 and samples in A.7.

method \ROUGE	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
OracleA	0.23	0.3	0.25	0.09	0.13	0.1	0.22	0.28	0.22
OracleEA	0.23	0.29	0.25	0.09	0.12	0.1	0.21	0.27	0.21

TABLE 6.9: ROUGE scores for summaries generated using two oracle-based models.

The models from this experiment showed approximately the same results (among themselves): ROUGE-1 F, ROUGE-2 F are the same, and ROUGE-L F is 0.01 more for OracleEA than for OracleA. Therefore, we will make a comparison of other models with the best model for ROUGE in this experiment. Also, these models were trained on different numbers of steps: OracleA at 30,000 and OracleEA at 40,000, respectively. OracleEA and OracleA models showed better results compared to 6.5.1. The values of ROUGE-1 F, ROUGE-2 F, and ROUGE-L F are greater at 0.04.

6.5.6 BertSumAbsClean

We found out that generated news incorporated text that does not apply to the sports event; this text in common cases located at the end of the news. See section 3.3 for more details. In this experiment, we eliminate sentences with such text.

We deleted sentences that contained one of the following words:

"таблица"/"здесь."/"онлайн-трляцяц
"table"/"here."/"online broadcast"

in broadcasts (source sequence) as well as in the news (target sequence). In broadcasts, a sentence with these words advertises online broadcasts on this site. In the news, sentences that contained these words referred either to another page or to a visualization (images/table); this information did not help to generate news and increase input sequences. Often such sentences have advertised mobile applications. The results of BertSumAbsClean experiments shown in table 6.10 and examples in table A.2.

method \ROUGE	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
BertSumAbsClean	0.18	0.24	0.19	0.07	0.07	0.06	0.18	0.18	0.16

TABLE 6.10: ROUGE scores for summaries generated from BertSum-AbsClean model.

We have noticed that metric ROUGE decreased compared to previous experiments, and we got the best model by ROUGE, the trained model only 20000 steps (this is the lowest number of training steps in our experiments). We compare the results with 6.5.1 because OracleA and OracleEA are experimental models. The ROUGE-1 F and ROUGE-L F metrics are less than 0.02, and ROUGE-2 F is less than 0.004. We hypothesis that deleted sentences were increasing our ROUGE: "gold" and generated news had advertisements and "referred's" sentences, and they increase the ROUGE.

6.5.7 AugAbs

According to previous experiments, the best models on ROUGE scores are based on neural networks. There are several works (Nallapati et al., 2016; Gavrilov, Kalaidin,

and Malykh, 2019; Tan, Wan, and Xiao, 2017b) show good results on relatively large amounts of data, and weak on small ones. Our dataset contained near 8000 data (chapter 3 more describe out data corpus). In the current experiment, we decided to increase our data corpus using augmentation technology. For this experiment, we decided to increase our data by ten times using synthetically generated data based on existing ones.
Our models for Augmentation are based on the work (Zhang, Zhao, and LeCun, 2015). The idea is to replace words in a broadcast on the words from another model. There are two models: thesaurus and static embedding. Both models receive a word at the input and return a list of words size of 10. Each word in the set is similar to the incoming word: the more similar words, the higher it in the returned list. Next, we use a geometric distribution to select two parameters for the model(for each generated sample): the number of words to be replaced in the broadcast and the index of word in the returned list for each word, that should be replaced.
Next, we describe two models. For the first model, called AugAbsTh, we used a similarity graph model of words from a Russian language thesaurus project called Russian Distributional Thesaurus [13]. The word similarity graph is a distribution thesaurus for the most frequent words of the Russian language, obtained on the embedding of words, which was built on the body of texts of Russian books (12.9 billion words).
For the second model, called AugAbsW2V, we used word2vec (Mikolov et al., 2013b) for vectorizing words and the cosine of the angle between the vectors, as a metric for word similarity. As a pre-trained model, we used a model trained on the Russian National Corpus (Kuratov and Arkhipov, 2019).
Since there were several news items related to one broadcast in our dataset, we did not augment the news. We have set random news that was written after a sports game; for broadcasts with less than ten news, we repeated the news. Thus, we got two datasets with sizes of about 80,000 broadcasts. We used the best model from previous experiments (6.5.4) for current data. The results of this and all previous models can be seen in table 6.11.

method \ *ROUGE*	*ROUGE-1*			*ROUGE-2*			*ROUGE-L*		
	P	R	F	P	R	F	P	R	F
Oracle	0.2	0.22	0.21	0.02	0.02	0.02	0.18	0.2	0.19
OracleA	0.23	0.3	0.25	0.09	0.13	0.1	0.22	0.28	0.22
OracleEA	0.23	0.29	0.25	0.09	0.12	0.1	0.21	0.27	0.21
PageRank W2V	0.06	0.15	0.08	0.	0.	0.	0.06	0.15	0.06
PageRank FT	0.06	0.13	0.08	0.	0.	0.	0.06	0.13	0.06
TextRank Gns	0.05	0.17	0.08	0.	0.01	0.	0.05	0.16	0.06
LexRank	0.07	0.1	0.08	0.	0.	0.	0.07	0.09	0.06
BertSumAbs	0.19	0.25	0.21	0.07	0.1	0.07	0.17	0.23	0.18
RuBertSumAbs	0.14	0.26	0.17	0.04	0.1	0.06	0.13	0.24	0.13
BertSumExtAbs	0.18	0.25	0.2	0.06	0.1	0.07	0.17	0.23	0.17
BertSumAbs1024	0.2	0.25	0.21	0.07	0.1	0.08	0.18	0.23	0.18
AugAbsTh	0.26	0.3	**0.26**	0.12	0.14	**0.13**	0.23	0.28	**0.25**
AugAbsW2V	0.23	0.26	0.22	0.08	0.1	0.09	0.19	0.25	0.21
BertSumAbsClean	0.18	0.24	0.19	0.07	0.07	0.06	0.18	0.18	0.16

TABLE 6.11: ROUGE scores from all models.

[13] https://nlpub.mipt.ru/Russian_Distributional_Thesaurus

Both models from this experiment indicated significant improve performance of our task and was training on 100000 steps. However, the AugAbsTh model showed a higher ROUGE score than the AugAbsW2V: the score of ROUGE-1 F, ROUGE-2 F, and ROUGE-L F are 0.04 more. This indicates that using synonyms to generate words in our task is more robust and significantly better than using word2vec embedding. AugAbsTh model has outperformed the best previous model BertSumAbs1024 as well as the oracle models. The score of ROUGE-1 F and ROUGE-2 F are higher by 0.5 and ROUGE-L F scores higher by 0.7 compared to BertSumAbs1024. Comparing with the OracleA model, AugAbsTh has ROUGE-1 F score higher on 0.1, ROUGE-2 F on 0.03, and ROUGE-L on 0.04 accordantly. This suggests that increasing the data corpus using real or "similar to real" data will increase the performance of the models. Table A.8 shows samples generated by the AugAbsTh model.

6.6 Human evaluation

In this section, we make a human evaluation for some results and analyze the coherence and consistency capacity of our model. We analyze the ability model to detect main events, competition teams, participants, and results in the sports game. We use 2 samples for evaluation presented in table A.9 and table A.10. In order to avoid misunderstandings, in this chapter, we will use the following notation: the broadcast is the incoming sequence, the "gold" news is the news written by a human, and the candidate news is the news that our model generated.

In the first example presented in table A.9, the model shows the coach interview after the match, in which the coach describes the game and results. Jürgen Klopp is the coach of the Borussia Dortmund (he was during the match), but the broadcast does not contain this information about Jurgen Klopp. The model also accurately determined teams in the game: the Borussia Dortmund football club versus FC Schalke. Gold news describes the results of the competition and the player's behavior after the game. However, the candidate news talks about comments from the head coach. The model indicates well which team won, but incorrect indicate the final score. Also, there is no contradiction in the candidate news: the coach positively comments the game that leads to victory. Also, the model mentions the name of the coach at the beginning and end of the news; this suggests that the model can remember information and carries it throughout the entire output sequence. Also, in candidate news, there is an expression: "18th round of the Bundesliga" - this information is missing as in the news as in the broadcast, although the Bundesliga is a football league in Germany and Borussia Dortmund and FC Schalke are German.

In the following comparison, we use an example from table A.10. We notice that "gold" and candidate news also differ in the type of storytelling, like in the previous: gold news describes main events in the current game and results of previous games; the news mainly describes the results of the goalkeeper Ilya. The candidate news looks in the format of an interview from the head coach of "Ak Bars" Zinetul Bilyaletdinov. The model successfully identifies the teams in the competition as well as the type of game, rules, and league: the candidate news use such related words as "puck" and the Kontinental Hockey League (KHL). The incoming sequence does not contain information about the trainer; the model received this information from previous broadcasts and news (during the training stage). Also, in the candidate news, we see the inconsistency in the fact that "Ak Bars" won against CSKA (according to the coach comments). At first, the coach says they won with score 3:1, but at the end of the interview, he contradicts those words, saying that the opponent created many moments, and their team missed two goals.

6.7 Compare human judgment and score ROUGE

The effectiveness of ROUGE was previously evaluated (Lin, 2004; Graham, 2015) through statistical correlations with human judgment on the DUC datasets (Paul Over, 2001; Paul Over, 2002; Paul Over, 2003). However, in these works, the authors used a different dataset, models, and environment than in the current work. In this section, we compare scores of the ROUGE metric and human judgments for random news from abstractive models. To judge the news, we asked five annotators to rate the news by five dimensions: relevance (selection of valuable content from the source), consistency (factual alignment between the summary and the source), fluency (quality of individual sentences), and coherence (collective quality of all sentences). We chose five random news from different models (with different number of training steps). The summary score for each dimension obtained by averaging the individual scores. In this task, we did not set the goal of statistically proving or disproving the correlations between the score ROUGE and human judgment. We want to get an estimate of the generated news from annotators, and visually compare how similar this rating is to score ROUGE. The comparison results are displayed in table 6.12. Analyzing the data from table 6.12, we want to emphasize that the values of Fluency and Coherence are generally higher than the values of Relevance and Consistency. This suggests that the models from our experiments generate pretty high-quality and linked sentences, but worse select events from the broadcast. The highest scores of Fluency and Coherence own the news #3 and #5; however, their score ROUGE-1 F differs by 6. News #1 #2 and #3 have the highest scores of Relevance; however, the scores of ROUGE-1 F it is also different: 0.19, 0.21, and 0.25, respectively. Concluding this experiment, we did not observe any visual relationships between human judgment and the ROUGE metric. We also want to notice that we received some comments from annotators regarding the quality of the news. Most of the comments were aimed at the fact that the quality of the sentences is pretty good, but the news does not review important events or reviews non-existent events from the broadcast.

dimention \score	*HJS*	*ROUGE-1*	*ROUGE-2*	*ROUGE-L*
news #1, BertSumAbs1024, 10000 steps				
Relevance	0.36			
Consistency	0.58	0.19	0.06	0.15
Fluency	0.56			
Coherence	0.46			
news #2, BertSumAbs1024, 30000 steps				
Relevance	0.26			
Consistency	0.34	0.21	0.08	0.18
Fluency	0.7			
Coherence	0.54			
news #3, OracleA, 40000 steps				
Relevance	0.46			
Consistency	0.6	0.25	0.1	0.22
Fluency	0.78			
Coherence	0.78			
news #4, BertSumAbs, 50000 steps				
Relevance	0.28			
Consistency	0.5	0.21	0.07	0.18
Fluency	0.56			
Coherence	0.58			
news #5, BertSumAbsClean, 20000 steps				
Relevance	0.28			
	0.56	0.19	0.06	0.16
Fluency	0.72			
Coherence	0.7			

TABLE 6.12: Comparison between human annotators and ROUGE scores for 5 news along different dimensions and models. HJS - human judgment score.

Chapter 7

Conclusion

In the current chapter, we will summarize our work, discuss results, and talk about future directions. In this paper, we investigated the problem of generating news based on sports commentary, and this task was tackled as the problem of summarization. Although we used SOTA approaches, we did not get the declared results. We assume there are several reasons for this. First of all, we would like to emphasize that these datasets use the English datasets, not Russian. These datasets contain significantly more documents than our case (Hermann et al., 2015b; Sandhaus, 2008; Narayan, Cohen, and Lapata, 2018), and the average size of one document is much larger in our case than in the case of the SOTA algorithms used. Last, we would like to note that the nature of the collection of buildings data corpus is also different. CNN/DM/XSumm datasets consist of public news and articles as input documents. The output summaries for them can contain either small summaries in one sentence (first sentence or news headline), summaries written by the same person after reading the input news or summaries which are obtained by the heuristic algorithmic way (Nallapati et al., 2016). Our dataset, as we showed in chapter 3, contains broadcasts at the input and news as output sequences, written by different people in a different context.
We also want to emphasize that there are fewer works that explored our problem. Existing works partially solved this problem, some of which generated sports news from short and robust descriptions of events in the game; others generate news from other news.

7.1 Contribution

In this work we made the following contribution:

- We investigate the application of the extractive approach to tackle our challenge and have got the maximum ROUGE score is 0.21 F1 score, utilizing the Oracle model described in Chapter 6.3. We found out that TextRank and LexRank models are not suitable for this task; their F1 score is 0.08. This suggests that solving this problem in an extractive way will not give significant results by the ROUGE metric.

- We found that the NMT approach gives the worst results if RNN-based models are used as an encoder. We assume that this is due to the larger size of the input sequence and the fact that RNN-based models do not work well with this property.

- However, if transformers are used as an encoder in the NMT approach, this improves the performance. We were reaching the maximum value 0.26 by

ROUGE-1F score using BERT as a transformer. We also notice that the multilingual BERT model showed better results for the Russian language, in contrast to the Russian-language RuBERT. We draw a conclusion that future research in this problem should be directed precisely using abstractive approaches.

- We also obtain the evidence of conclusions that Tan has made in previous works (Liu and Lapata, 2019; Tan, Wan, and Xiao, 2017b). Tan and other concluded that the combination of extractive and abstractive approaches improves the performance of summarization tasks. We got a 0.25 F1 score when using the Oracle model in the first stage, and the encoder-decoder model in the second.

- We found out that increasing data corpus using text argumentation based on thesaurus results in a substantial improvement: we increase data per ten times, and the ROUGE-1 F score has up on 0.5 in absolute difference. We have shown it in Chapter 6.5.7.

- And in the end, we would like to emphasize that the ROUGE metric is not suitable for the metric in our problem. Analyzing some of the results, we found out that the generated news did not consider the events and results of the game. However, they have a high score of ROUGE and were logical in the structure (chapter 6.6 shows a detailed analysis of several examples).

7.2 Future work

We have several directions for future work:

- In Chapter 6.5.7, we showed an improvement in performance by increasing the volume of data corpus; we will move further in this direction. We plan to engage other companies used the text broadcasting of sporting events. We also plan to transform the comments of the sporting event from audio sources, which are more popular than textual.

- The effective application of transformers as an encoder demonstrated in Chapter 5 suggests that we will experiments with other types of transformers, like GPT-2 or BERT with other parameters in the future.

- We leveraged the ROUGE metric in the current work, which based on the overlapping of n-gram between candidate and gold summary. We want to use a metric that based on detected main characteristics of the game: overall score, teams in the competition, key people, main events, the conclusion of the game. This is the topic of our future research.

Appendix A

Examples of generated news

нападающии цска дарси веро подрался с форвардом ска евгением артюхиным на 2 - и минуте матча .оба игрока удалены на пять минут .кроме того , веро получил две минуты за удар клюшкои. sports . ru ведет текстовую трансляцию матча цска [UNK] ска .видео онлаин - трансляцию матча вы можете посмотреть здесь .начало [UNK] в 19. 00 по московскому времени .таблица хлстатистика хлтаблица хл sports . "CSKA" striker Darcy Vero had a fight with striker SK Evgeny Artyukhin at the 2 - minute of the match. Both players were removed for five minutes. In addition, Vero received two minutes for a club hit. sports. ru conducts the text broadcast of the match CSKA [UNK] ska. video onlain - you can watch the broadcast of the match here. start [UNK] at 19. 00 Moscow time. khlstatistics table khl khl table sports.
полузащитник « краснодара » вандерсон поделился впечатлениями от матча 12 - го тура премьер - лиги против цска (1 : 0) . « очень сложныи матч .мы создали хорошие моменты , но не реализовали их .к сожалению , не забили два гола .для меня это очень важная победа » , [UNK] сказал фернандес. sports . ru проводил онлаин - трансляцию этого матча .начало [UNK] в 19 . 00 по московскому времени .таблица чемпионата россиистатистика чемпионата россии sports .таблица рфплтаблица премьер "Krasnodar" midfielder Wanderson shared his impressions of the match of the 12th round of the Premier League against "CSKA" (1: 0). "A very difficult match. We created good moments but did not realize them. Unfortunately, we did not score two goals. For me it is a very important victory," [UNK] Fernandez said. sports.ru conducted online - broadcast of this match. First [UNK] at 19. 00 Moscow time .the championship table of Russia statistics of the championship of Russia sports .the table of the Russian Premier League the table of the Premier League sports the table of the Premier

TABLE A.1: Examples of summaries generated using BERTABS model.

Comments

в турине дортмундцы потеряли не только физические и моральные силы, но и пищека с папастатопулосом. в центре обороны сегодня будут деиствовать хуммельс и суботич. а на фланге сыграет кирх, для которого это лишь второе попадание в стартовыи состав на матч бундеслиги по ходу сезона. привет фанатам бундеслиги ! юрген клопп пару - троику лет назад сказал, что в календаре чемпионата есть тридцать две обычных игры и два дерби и данныи факт чуть ли не прописан в его контракте с клубом. возможно, немец просто утрировал, но сегодня мы станем свидетелями одного из самых принципиальных матчеи в европеиском футболе : боруссия против шальке, дортмунд против гельзенкирхена, шварцгельбен против кенигсблауэн. у шальке тоже есть потери, и средь них по - прежнему выделяеця дракслер. но, поскольку команда вынуждена обходиться без юлиана не первыи месяц, все уже вроде и привыкли. ворота продолжает защищать девятнадцатилетнии велленроитер, хотя в запасе остаеця опытныи веткло, известныи по выступлениям за маинц. принимая во внимание расписания команд, можно заметить, что на стороне шальке несправедливое преимущество. команда роберто ди маттео в прошлую субботу сыграла с вердером и целую неделю могла готовиться к дерби. боруссия же вечером вторника билась с ювентусом у подножия альп, после чего времени в ее распоряжении оставалось гораздо меньше. разумееця, помимо текстовои онлаин - трансляции вашему вниманию будут предложены видеофрагменты самых ярких момент

in Turin, the Dortmund lost not only their physical and moral strength but also the Pisces with papastatopoulos. Hummels and Subotic will be at the center of defense today. and on the flank will play a Kirch, for which this is only the second hit in the starting lineup for the Bundesliga match during the season. hello to the Bundesliga fans! Jürgen Klopp said a couple of years ago that there were thirty-two regular games and two derbies in the championship calendar, and this fact was almost spelled out in his contract with the club. the German may have simply exaggerated, but today we will witness one of the most important matches in European football: Borussia against Schalke, Dortmund against Gelsenkirchen, Schwarzgelben against Kenigsblowen. the Schalke also has losses, and among them, the Draxler still stands out. but, since the team is forced to do without Julian for several months, everyone seems to be used to it. the gate continues to be defended by a nineteen-year-old vellenroiter, although the experienced veto remains in reserve, known for playing for the Mainz. taking into account the team schedules, you can see that the Schalke side has an unfair advantage. Roberto di Matteo's team last weekend playing with the Werder Bremen and could spend a week preparing for the derby. Borussia, on Tuesday evening, fought with Juventus at the foot of the Alps, after which time it had much less time at its disposal. Of course, in addition to text online broadcasts, video clips of the most striking moments will be offered to your attention

Candidate news главный тренер дортмундскои ”боруссии” юрген клопп поделился впечатлениями от матча 18 - го тура бундеслиги против «шальке» (1 : 0). ”мы очень довольны .боруссия заслужила эту победу. мы очень хорошо играли в первом таиме, но во втором смогли переломить ход встречи. во втором тайме мы действовали неплохо, но не смогли выйти в атаку”, [UNK] сказал клопп. The head coach of the Dortmund Borussia Jürgen Klopp shared his impressions of the match of the 18th round of the Bundesliga against Schalke (1: 0). "We are very pleased. Borussia deserved this victory. We played very well in the first game, but in the second we were able to turn the tide of the meeting. in the second game, we acted well, but we couldn't attack, "[UNK] said, Klopp.
"Gold" news ”боруссия” дортмунд ведет в матче чемпионата германии против ”шальке” (2:0 , второй тайм). нападающий хозяев пьер-эмерик обамеянг отпраздновал первый гол в матче с форвардом марко ройсом. габонец надел на себя маску бэтмена, а ройс – маску робина. "Borussia" Dortmund leads in the match of the championship of Germany against "Schalke" (2: 0, second half). home striker Pierre Emerick Obameyang celebrated the first goal of the match with striker Marco Royce. The Gabonets put on a batman mask and Royce a robin mask.

TABLE A.9: Table show components of test example 1 for the human evaluation.

Comments

привет всем, кто не представляет своеи жизни без хоккея ! сегодня маркетинговое реалити - шоу, призванное улучшить посещаемость матчеи ак барса, вряд ли потребуеця . ведь в казань приезжают великии и ужасныи ’ ’ александр радулов и павел дацюк, выступавшии за хозяев в сезоне 2000 / 01 . интересно, что последняя встреча команд в казани состоялась почти год назад - 18 - го октября . тогда дубль иммонена принес хозяевам победу . но тогда у цска не было ни братьев радуловых, ни дацюка, ни грабовского, ни брызгалова . . . в хл команды провели 8 матчеи, 6 раз победа оставалась за казанцами . но нынешние главные тренеры белов и брагин ранее в очном противостоянии не встречались . лучшими бомбардирами команд являюця : у ак барса ’ ’ - алексеи терещенко (3 Ѣ 9) ; у цска - александр радулов (7 Ѣ 12) . стартовые вратари : константин барулин - илья брызгалов . понеслась ! зарипов предпринял попытку прорваться вперед на скорости - не успел к шаибе . обстановка у ворот хозяев нагнетаеця . зарипов - 1 : 0 ! ! ! 2 - и гол в сезоне ! контратака, морозов получил шаибу в центре, сместился вправо, сделал диагональ на дальнюю штангу - зарипов замкнул ! все армеицы не успели на один шаг ’ ’ в даннои ситуации . чрезвычаино насыщен голевыми шансами матч с самого начала ! отбор дацюка, пас в центр на квашу, тот красиво уходит от соперника и . . . совсем плохо бросает с близкои дистанции . зарипов сегодня вырвался на площадку как та самая

hello to all who cannot imagine their life without hockey! Today, a marketing reality show designed to improve attendance at Ak Bars matches is unlikely to be required. after all, the great and terrible come to Kazan ", Alexander Radulov and Pavel Datsyuk, who played for the hosts in the 2000/01 season. It is interesting that the last team meeting in Kazan took place almost a year ago - on October 18th. then the double of Immonen brought the hosts a victory. but then the Tssk had no Radulov brothers, no Datsyuk, no Grabovsky, no Squirrel. . . in the khl team spent 8 matches, 6 times the victory remained for Kazan. but the current head coaches of Belov and Bragin had never met before in full-time confrontation. The best scorers of the teams are: Ak Bars " - Alexei Tereshchenko (3 + 9); CSKA - Alexander Radulov (7 + 12). starting goalkeepers: Konstantin Barulin - Ilya squirting. raced! Zaripov made an attempt to break forward at speed - did not have time to shaib. the situation at the gates of the hosts is being pumped up. Zarypov - 1: 0! ! ! 2 - and the goal of the season! counterattack, frost got a shaib in the center, shifted to the right, made a diagonal to the far post - he closed it! all the armies did not have one-step " in this situation. Perekrestichino is full of chances to score the match from the very beginning! selection of Datsyuk, passing to the center for Kvash, that beautifully leaves the opponent and. . . very badly throws from close range. Zarypov today burst onto the site as the same

Candidate news

главныи тренер ”ак барса” зинэтула билялетдинов поделился впечатлениями от победы над цска (3 : 1) в матче регулярного чемпионата хл. ”игра была очень интересная, боевая, боевая. обе команды создали много моментов, но соперник создал больше моментов. во втором периоде мы пропустили две шайбы. но, в целом, я доволен и результатом”, [UNK] цитирует билялетдинова ”спорт - экспресс”.

head coach of "Ak Bars" Zinetula Bilyaletdinov shared his impressions of the victory over CSKA (3: 1) in the match of the regular championship of KhL. "the game was very interesting, fighting, fighting. both teams created a lot of moments, but the opponent created more moments. in the second period we missed two goals, but, in general, I am pleased with the result," [UNK] quotes Bilyaletdinov “sport - express "

"Gold" news

голкипер цска илья брызгалов пропустил по 4 шайбы в трех сыгранных матчах в хл из четырех.вратарь « филадельфии », выступающий за цска во время локаута, четыре раза вынимал шайбу из своих ворот в матчах с «трактором » (3:4), ска (2:4) и « ак барсом » (2:5).«ч»пятая шайба в матче с « ак барсом » была заброшена в пустые ворота.в единственном победном матче за цска (против « салавата юлаева » – 4:1) брызгалов пропустил одну шайбу.в этих играх коэффициент надежности брызгалова составил 3,27, а процент отраженных бросков – 88,7%. подробную статистику брызгалова смотрите здесь.

CSKA goalkeeper Ilya Pryzgalov conceded 4 goals in three out of four games played in KHL. The Philadelphia goalkeeper, who played for CSKA during the lockout, took the puck out of his goal four times in matches with Tractor (3: 4), ska (2: 4) and “Ak Bars” (2: 5). <q> The fifth goal in the match with “Ak Bars” was scored into the empty net. In the only winning match for CSKA (against “Salavat Yulaev” - 4: 1) Skrygalov missed one goal. In these games, Kryzgalov’s reliability coefficient was 3.27, and the percentage of reflected shots was 88.7%. see detailed statistics here.

TABLE A.10: Table show components of test example 2 for the human evaluation. This example includes broadcast information, using as the input source, candidate news, generated by the model, and "gold" news, written by a human.

нападающии « челси » фернандо торрес забил два гола в матче 6 - го тура группового этапа лиги чемпионов с « вильярреалом » (2 : 0 , первыи таим) .таким образом , теишеира забил 22 - и гол в составе « горожан » . sports .ru вел текстовую трансляцию этого матча Chelsea striker Fernando Torres scored two goals in the 6th round match of the Champions League group stage with Villarreal (2: 0, first). Thus, Teesheira scored 22 - and a goal as part of the townspeople .sports .ru led the text broadcast of this match
нападающии магнитогорского « металлурга » сергеи мозякин в матче с « авангардом » забросил первую шаибу в хл .форвард отличился в середине второго периода , сделав счет 2 : 1 в пользу питерского клуба .для мозякин стал 306 - м в регулярных чемпионатах хл и матч .всего в нынешнем сезоне мозякин провел 99 матчеи в которых набрал 112 (50 62) очков при показателе полезности 26 .п . с . « виртуальныи букмекер » на sports . The striker Magnitogorsk “metallurgist” Sergei Mozyakin in a match with “Vanguard” threw the first Shiba in the KHL forward. He scored in the middle of the second period, making the score 2: 1 in favor of the St. Petersburg club. For Mozyakin he became 306th in the regular championships of the KHL and the match. This season Mozyakin played 99 matches in which he scored 112 (50 62) points with a utility score of 26 .p. s. "Virtual bookmaker" at sports.

TABLE A.2: Examples of summaries generated using BertSumOracletAbs and BertSumOracleExtAbs models. They showed similar results, so we do not divide samples of summaries by models.

PageRank+word2vec: **Очень точно заметил Юрий Жирков о том, что футболистам ЦСКА было абсолютно все равно, с кем играть, ведь если ставить самые высокие цели, то надо быть готовыми попасть на третью команду английской премьер-лиги (текущее место) уже на стадии 1/16 финала.** Эшли Янг ворвался в штрафную площадь, упал, надеясь на пенальти, но главный арбитр занял правильную позицию и показывает, что игроку "Астон Виллы" нужно подниматься. **Yuri Zhirkov noted that the "CSKA" players did not care who to play with, because if you set the highest goals, you need to be ready to get on the third team of the English Premier League (current place) already at the 1/16 final stage.** Ashley Young burst into the penalty area, fell, hoping for a penalty. Still, the chief referee took the right position and shows that the player, "Aston Villa" needs to rise.
PageRank+FastText: Вагнер пяткой отдал мяч на Дзагоева, тот вывел Алексея Березуцкого почти один на один по центру, но защитник пробил очень плохо, направив мяч много правее штанги! Алдонин покинул поле, Бэрри очень грубо въехал ему по ногам, но арбитр не удостоил того даже предупреждением! **Очень точно заметил Юрий Жирков о том, что футболистам ЦСКА было абсолютно все равно, с кем играть, ведь если ставить самые высокие цели, то надо быть готовыми попасть на третью команду английской премьер-лиги (текущее место) уже на стадии 1/16 финала.** Wagner put ball to Dzagoev with his heel, he brought Aleksey Berezutsky almost one on one in the center, but the defender struck very badly, directing the ball much to the right of the bar! Aldon left the field, Barry very roughly ran into his legs, but the referee did not even dignify that with a warning! **Yuri Zhirkov noted that the "CSKA" players did not care who to play with, because if you set the highest goals, you need to be ready to get on the third team of the English Premier League (current place) already at the 1/16 final stage.**
Gensim TextRank: **Очень точно заметил Юрий Жирков о том, что футболистам ЦСКА было абсолютно все равно, с кем играть, ведь если ставить самые высокие цели, то надо быть готовыми попасть на третью команду английской премьер-лиги (текущее место) уже на стадии 1/16 финала.** Хочеця верить в то, что Вагнер Лав не просто продолжит забивать, но и установит личное достижение, побив вечный рекорд Юргена Клинсманна, забившего в рамках Кубка УЕФА сезона - 1995/96 отличился в составе "Баварии" 15 раз и помог мюнхенцам завоевать почетный трофей. **Yuri Zhirkov noted that the "CSKA" players did not care who to play with, because if you set the highest goals, you need to be ready to get on the third team of the English Premier League (current place) already at the 1/16 final stage.** I want to believe that Vágner Love will not only continue to score but also set a personal achievement, breaking the eternal record of Jürgen Klinsmann, who scored in the UEFA Cup of the season - 1995/96, scored 15 times in the Bayern Munich and helped the Munich team win the honorary trophy.

TABLE A.3: Examples of summaries generated using extractive approaches: PageRank+word2vec, PageRank+fastText + Gensim TextRank.

Сегодня у нас Лига Европы. И для того же “Андерлехт” матч сегодня принципиальнейший. Чуть-чуть не попал Константин в ворота, обидно. Все-таки индивидуальное и командное преимущество “Зенит” ощущаеця очень четко . 2:0. думаю , игра сделана .Прото спасает команду после удара Канунникова в упор ! 3:1 - все как в первом матче. Today we have the Europa League. And for the same “Anderlecht” match today is crucial. I almost got Konstantin through the gate, it's a shame. Still, the individual and team advantage of “Zenith” is felt very clearly. 2-0. I think the game is done. Proto saves the team after hitting Kanunnikov point-blank! 3: 1 - everything is like in the first match.
Состав испанцев предсказуем до неприличия , что не может не радовать поклонников этой команды . Все опасные моменты и голы матча можно будет посмотреть в нашем блоге поехали. Ватч начался! Иньеста врываеця в штрафную, прострел на Торреса. В каждой атаке он стараеця быть активным. Блистательный тайм в исполнении Парагвая! Но и сам Андрес немного замешкался. Замены явно зрели, испанцам нужны голы. Редкое для испанской сборной событие на этом турнире. Победа в этом матче станет для испании историческим событием. The composition of the Spaniards is indecently predictable, which cannot but please the fans of this team. All the dangerous moments and goals of the match can be seen in our blog! The match has begun! Iniesta breaks into the penalty area, backache on Torres. In every attack, he tries to be active. A brilliant half by Paraguay! But Andres himself hesitated a little. Substitutions clearly matured, the Spaniards need goals. A rare event for the Spanish national team in this tournament. Victory in this match will be a historic event for Spain.
В большинстве матчей ЦСКА решает свои проблемы задолго до конца , однако “Зенит” и “Локо” удивительно синхронно не дают сбоев .”Амкар” активнее и интереснее провел начало матча . Плотная игра в центр . А что , очень даже эстетичн , с африканской такой пластикой. Фернандесу досталос , только непонятно, как. Потому что раскрепощенный ”Амкар” дома бы , скорее всего , забил. Защитник считал, что он обогнал бы Думбия в любом случае. Канунников теперь будет напрягать оборону ЦСКА в центре . ”Амкар” вряд ли можно в чем-то себя упрекать. In most matches, CSKA solves their problems long before the end, however, “Zenith” and “Loko” surprisingly synchronously do not fail. “Amkar” more actively and interestingly started the match. Intense game in the center. And that, very aesthetic, with such African plastic. Fernandez got tired, only it is not clear how. Because the liberated “Amkar” at home would most likely have scored. The defender believed that he would have overtaken Doumbia in any case. Kanunnikov will now strain the defense of CSKA in the center. “Amkar” can hardly be blamed for something.

TABLE A.4: Examples of summaries generated using Oracle model.

полузащитник « спартака » деми де зеув прокомментировал результат матча 10 - го тура премьер - лиги против « амкара » (1 : 1) . « мы довольны результатом , потому что у нас были моменты .надо отдать должное « амкару » .они хорошо отыграли в атаке , но нам не хватило инициативу и пропустили на контратаках .в целом , я думаю , что победили » , цитирует збоа « спорт - экспресс » . подробную статистику матча смотрите здесь .« спартак » [UNK] « амкар » вы можете посмотреть здесь . амкар .онлаин sports . ru провел онлаин - трансляцию этого матча .начало [UNK] в 17 . 00 по московскому времени .таблица премьер - трансляцию этои игры .статистика красно - лигистатистика премьер - трансляциистатистика крылья советов » видео таблица краснодарь россии « амкаром »статистика красных турниров красно - лигеподробнаяцию матча вы можете проплстатистика рфплвидео Spartak midfielder Demi de Zeuv commented on the result of the match of the 10th round of the Premier League against Amkar (1: 1). "We are happy with the result, because we had moments. We must pay tribute to the Amkar. They played well in the attack, but we lacked the initiative and missed on the counterattacks. In general, I think we won," sports - quotes zboa express. " see detailed statistics of the match here. Spartak [UNK] Amkar you can see here. amkar .onlain sports. ru spent onlain - broadcast of this match. the beginning [UNK] at 17. 00 Moscow time. Prime table - broadcasting these games. Statistics red - linguistic statistics Prime - broadcasting statistics wings of soviets »video table Krasnodar Russia" Amkar "statistics of red tournaments red - detailed match statistics you can play statistics rfplvideo.
нападающии ска алексеи яшин стал лучшим игроком выставочного матча против « каролины » (4 : 3) .на счету 36 - летнего игрока в этои встрече гол и две результативные передачи . 18 - летнии форвард отметился дублем . в нынешнем сезоне на счету ковальчука 4 (2 1) очка в 18 играх .подробную статистику выступлении игрока смотрите здесь . sports . ru провел текстовую трансляцию матча .начало [UNK] в 19 . 00 по московскому времени .соберите свою команду увесия » , [UNK] цитирует хоккеиста можно посмотреть здесь .fantasy hockey .сочи - 2014 [UNK] соберите события собери : финляндия [UNK] собери : б .соберите события .уведомлением ! соберите текстовыи онлаин этого матча вы можете по рссяндекс - трансляцию этои встречи . [УНК] вы можете прогно - виджете посмотреть здесьа вы можете прочитать здесь . п . с .« виртуальныи букмекер » на ставках без риска новыи год в лондоне в качестве главного приза .слухи : ваша команда мечты !« трибуна » на striker ska alexei yashin became the best player of the exhibition match against "carolina" (4: 3). On the account of the 36-year-old player in this meeting a goal and two assists. An 18-year-old striker scored twice. this season, Kovalchuk has 4 (2 1) points in 18 games. For detailed statistics on the player's performance, see here. sports. ru conducted a text translation of the match. First [UNK] at 19. 00 Moscow time. Assemble your weight team ", [UNK] quotes from the hockey player can be seen here .fantasy hockey Sochi - 2014 [UNK] collect events collect: Finland [UNK] collect: b. collect events. You can collect textual onlain of this match by rssyandeks - broadcast this meeting. [UNK] you can forecast - you can see here, you can read here. p. s. "virtual bookmaker" at risk-free bets for the new year in London as the main prize. rumors: your dream team! "tribune" on

TABLE A.5: Examples of summaries generated using RuBERTABS model.

исполняющии обязанности главного тренера челси роберто ди маттео поделился впечатлениями от матча 27 - го тура чемпионата англии с уиганом (1 : 0) . я очень доволен .эта ничья была очень интересная игра .в концовке первого таима мы создали достаточно моментов , но не смогли реализовать свои моменты .а в целом результатом я доволен ничьеи с уиганом .но я доволен и игрои своеи игрои .я доволен ничью с уиганом деиствовал очень доволен победои , приводит слова ди маттео bbc .напомним , что матч с уиганом Roberto di Matteo, acting head coach of Chelsea, shared his impressions of the match of the 27th round of the championship of England with wigan (1: 0). I am very pleased. This draw was a very interesting game. At the end of the first period, we created enough moments, but could not realize our moments. On the whole, I am satisfied with a draw with Wigan. But I am satisfied with my games and my games. I am satisfied with a draw with Wigan acted very pleased with the victories, leads the words di Matteo BBC. recall that the match with Wigan
форвард « анжи » диего тарделли дебютировал в составе своеи новои команды в матче 18 - го тура премьер - лиги против « манчестер сити » (2 : 1) . на 30 - и минуте он вышел на поле на поле вместо одила ахмедова .напомним , что махачкалинскии клуб пропустил в премьер - лиге из - за травмы .подробную статистику выступлении игрока смотрите здесь .sports . ru проводит онлаин - трансляцию этого матча .начало [UNK] в 21 . 00 по московскому времени .sports , игра в 16 . ru проведет онлаин - трансляции матча The forward Anji Diego Tardelli made his debut as part of his new team in the match of the 18th round of the Premier League against Manchester City (2: 1). at the 30th minute, he entered the field instead of Akhmedov. we recall that the club missed the Makhachkala club in the Premier League because of an injury. For detailed statistics on the player's performance, see .sports here. ru holds online - broadcast of this match. Start [UNK] at 21. 00 Moscow time .sports, the game is at 16. ru will conduct online - match broadcasts

TABLE A.6: Examples of summaries generated using BertSum-Abs1024 model.

главныи тренер сатурна анатолии баидачныи после поражения в матче 19 - го тура премьер - лиги с амкаром (0 : 1) заявил , что в этои встрече его команде не хватило грамотно .соперник создал больше моментов , чем в первом таиме .а во втором таиме , я думаю , игра была равнои , сказал тренер в эфире нтв - плюс .подробную статистику этого матча вы можете посмотреть здесь .победа была равная , сказал баидачныи в эфире телеканала спорт - экспресс .видеоподробнее о прошедшеи игре вы можете по Saturn's head coach Anatoly Baidachny after losing in the match of the 19th round of the Premier League with Amkar (0: 1) said that his team did not have enough competitors at this meeting. The rival created more moments than in the first game. and in the second period, I think the game was equal, the coach said on NTV - plus. You can see the detailed statistics of this match here. The victory was equal, said the best on the air of the sports - express television channel. You can learn more about the past game on
полузащитник « реала » криштиану роналду оформил хет - трик в первом таиме матча чемпионата испании против « хетафе » (4 : 0 , второи таим) . таким образом , 29 - летнии хавбек сборнои португалии забил 7 - и гол в составе « сливочных » . подробную статистику выступлении игрока смотрите здесь . sports . ru textcyr - « » [UNK] « реал » .начало [UNK] в 18 . 30 по московскому времени .спорц .таблица чемпионата испаниистатистика чемпионата испанииподробная статистика примерыподробная и таблица чемпионата испании Real Madrid midfielder Cristiano Ronaldo scored a hat-trick in the first half of the match of the championship of Spain against the "Getafe" (4: 0, second time). Thus, the 29-year-old midfielder of the Portuguese national team scored 7 - and a goal in the "cream". detailed statistics of the player's performance, see here. sports. ru conducts online - broadcast of the match "Real Sociedad" [UNK] "Real". The beginning of [UNK] at 18. 30 Moscow time .sports. Spanish championship table Spanish championship statistics detailed statistics example detailed and Spanish championship table

TABLE A.7: Examples of summaries generated using BertSumOracletAbs and OracleEA models. They showed similar results, so we do not divide samples of summaries by models.

главныи тренер цска дмитрии квартальнов прокомментировал победу над « магниткои » (2 : 1) в третьем матче финала кубка гагарина .[UNK] мы знали , что « магнитка » [УНК] очень организованная команда .в первом периоде у нас было много моментов , но мы их не использовали .[UNK] The head coach of CSKA Dmitry Quarterly commented on the victory over "Magnitogorsk" (2: 1) in the third match of the final of the Gagarin Cup. [UNK] we knew that "Magnitogorsk" [UNK] was a very organized team. In the first period, we had a lot of points, but we didn't use them. [UNK]
форвард « манчестер юнаитед » уэин руни , автор единственного гола в матче чемпионата англии против « ньюкасла » (1 : 0) , поделился впечатлениями об игре .« мы победили в тяжелои борьбе .нашеи задачеи было заработать три очка , и мы получили то , что хотели .у нас были моменты и в первом таиме , но мы их не реализовали .но знали , что моменты еще будут .к счастью , этот шанс выпал мне , и я забил .болельщики всегда ставят под сомнение боеспособность больших клубов , но мы Forward Manchester United Wayne Rooney, the author of the only goal in the match of the championship of England against the Newcastle (1: 0), shared his impressions of the game. "We won in a difficult fight. Our task was to earn three points, and we got what we like. we had moments for the first time, but we didn't realize them. but we knew that there would still be moments. fortunately, this chance fell on me, and I scored.
полузащитник « барселоны » и сборнои испании хави после матча квалификации евро - 2012 с чехиеи (2 : 1) заявил , что его команда заслуживала большего .« мы провели хорошии матч , владели инициативои , но соперник деиствовал уверенно .во втором таиме мы деиствовали получше , но в завершающеи стадии мы деиствовали не так , как хотелось бы » , [UNK] цитирует хави goal .com .отметим , что матч с чехиеи стал для испанцев четвертым в составе сборнои испании . the midfielder of Barcelona and the Spanish national team, Xavi, after the Euro 2012 qualification match with the Czech Republic (2: 1), said his team deserved more. "We had a good match, we had the initiative, but the opponent acted confidently. in the second time we acted better, but in the final stages, we didn't act as we would like, "[UNK] quoted goal.com Xavi. We note that the match with the Czechs was the fourth for the Spaniards in Spain.

TABLE A.8: Examples of summaries generated using AugAbsTh model.

главныи тренер « шахтера » мирча луческу прокомментировал победу над киевским « динамо » (2 : 0) в матче 25 - го тура чемпионата украины . « игра была очень тяжелая .тяжелая , что мы создали много опасных моментов , создали много моментов .во втором таиме мы не смогли реализовать численное преимущество .после перерыва мы стали деиствовать активнее » , [UNK] сказал луческу в эфире телеканала « футбол » .« шахтер » обыграл « черноморец » . sports . ru проводил текстовую онлаин - трансляцию этого матча . Shakhtar head coach Mircea Lucescu commented on the victory over Dynamo Kyiv (2-0) in the match of the 25th round of the Ukrainian championship. "The game was very difficult. It was difficult that we created many dangerous moments, created many moments. The second time, we could not realize the numerical advantage. After the break, we began to act more actively," [UNK] told Lucescu on the television channel Football. Shakhtar beat Chernomorets. sports.ru conducted a text online - broadcast of this match.
33 - летнии полузащитник « челси » фрэнк лэмпард досрочно покинул поле в матче 3 - го тура чемпионата англии против « норвича » (2 : 1) . в середине первого таима лэмпард получил травму и попросил замену .на 37 - и минуте вместо него вышел флоран малуда . подробную статистику выступлении игрока вы можете посмотреть здесь .sports . ru ведет онлаин - трансляцию матча « норвич » [UNK] « челя » .видеоонлаинначало [UNK] в 19 . 00 по московскому времени .таблица премьер - лигистатистика премьер - лиги Frank Lampard, the 33-year-old Chelsea midfielder, left the field ahead of schedule in the match of the 3rd round of the championship of England against the Norwich (2: 1). in the middle of the first time, the Lampard was injured and asked for a replacement. on 37 - and a minute a Floran of Malouda came out instead. You can see detailed statistics of the player's performance here .sports. ru leads online - broadcast of the match "Norwich" [UNK] "Chelsea." videoonlainstart of [UNK] at 19. 00 Moscow time. Premier League statistics table Premier League

TABLE A.11: Examples of summaries generated using BertSumExtAbs model.

Bibliography

A.I. Mikhailov A.I. Black, R.S. Gilyarevsky (1965). “Fundamentals of Scientific Information”. In: *The science*. (, 1965) , .. []/.. , .. , .. . –.: , 1965.–655 ., p. 655.

Ayala, Miguel Francisco Carmona (2019). “Generating Football Match Summaries with NMT”. In: URL: https://web.stanford.edu/class/cs224n/reports/custom/15723716.pdf.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2014). *Neural Machine Translation by Jointly Learning to Align and Translate*. arXiv: 1409.0473 [cs.CL].

Bojanowski, Piotr et al. (2016). “Enriching Word Vectors with Subword Information”. In: *CoRR* abs/1607.04606. arXiv: 1607.04606. URL: http://arxiv.org/abs/1607.04606.

Bouayad-Agha, Nadjet, Gerard Casamayor, and Leo Wanner (Sept. 2011). “Content selection from an ontology-based knowledge base for the generation of football summaries”. In: *Proceedings of the 13th European Workshop on Natural Language Generation*. Nancy, France: Association for Computational Linguistics, pp. 72–81. URL: https://www.aclweb.org/anthology/W11-2810.

Bouayad-Agha, Nadjet et al. (Aug. 2012). “Perspective-oriented Generation of Football Match Summaries: Old Tasks, New Challenges”. In: *ACM Trans. Speech Lang. Process.* 9.2, 3:1–3:31. ISSN: 1550-4875. DOI: 10.1145/2287710.2287711. URL: http://doi.acm.org/10.1145/2287710.2287711.

Celikyilmaz, Asli et al. (2018). *Deep Communicating Agents for Abstractive Summarization*. arXiv: 1803.10357 [cs.CL].

Chorowski, Jan K et al. (2015). “Attention-Based Models for Speech Recognition”. In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes et al. Curran Associates, Inc., pp. 577–585. URL: http://papers.nips.cc/paper/5847-attention-based-models-for-speech-recognition.pdf.

Chung, Junyoung et al. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. arXiv: 1412.3555 [cs.NE].

Cibils, André et al. (2018). *Diverse Beam Search for Increased Novelty in Abstractive Summarization*. arXiv: 1802.01457 [cs.CL].

Cohan, Arman et al. (2018). *A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents*. arXiv: 1804.05685 [cs.CL].

Culurciello, Eugenio (2019). *The fall of RNN / LSTM*. URL: https://towardsdatascience.com/the-fall-of-rnn-lstm-2d1594c74ce0.

Dang, Hoa (Jan. 2006). “Overview of DUC 2006”. In:

Devlin, Jacob et al. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805 [cs.CL].

eBizMBA Inc (2019). *Top 15 Most Popular Sports Websites: September 2019*. http://www.ebizmba.com/articles/sports-websites.

Erkan, Günes and Dragomir R. Radev (2011). “LexRank: Graph-based Lexical Centrality as Salience in Text Summarization”. In: *CoRR* abs/1109.2128. arXiv: 1109.2128. URL: http://arxiv.org/abs/1109.2128.

Gavrilov, Daniil, Pavel Kalaidin, and Valentin Malykh (2019). "Self-Attentive Model for Headline Generation". In: *CoRR* abs/1901.07786. arXiv: 1901.07786. URL: http://arxiv.org/abs/1901.07786.

Gehrmann, Sebastian, Yuntian Deng, and Alexander Rush (2018). "Bottom-Up Abstractive Summarization". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4098–4109.

Graefe, A., Columbia University. Graduate School of Journalism. Tow Center for Digital Journalism, and GitBook (2016). *Guide to Automated Journalism*. URL: https://books.google.com.ua/books?id=0iPbjwEACAAJ.

Graham, Yvette (Sept. 2015). "Re-evaluating Automatic Summarization with BLEU and 192 Shades of ROUGE". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, pp. 128–137. DOI: 10.18653/v1/D15-1013. URL: https://www.aclweb.org/anthology/D15-1013.

Gusev, Ilya (2019). *Importance of Copying Mechanism for News Headline Generation*. arXiv: 1904.11475 [cs.CL].

Harris, Zellig S. (1954). "Distributional Structure". In: *<i>WORD</i>* 10.2-3, pp. 146–162. DOI: 10.1080/00437956.1954.11659520. eprint: https://doi.org/10.1080/00437956.1954.11659520. URL: https://doi.org/10.1080/00437956.1954.11659520.

Helen, Afrida (2018). "Automatic Abstractive Summarization Task for New Article". In: *EMITTER International Journal of Engineering Technology* 6.1, 22–34. DOI: 10.24003/emitter.v6i1.212.

Hermann, Karl Moritz et al. (2015b). "Teaching Machines to Read and Comprehend". In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes et al. Curran Associates, Inc., pp. 1693–1701. URL: http://papers.nips.cc/paper/5945-teaching-machines-to-read-and-comprehend.pdf.

Hermann, Karl Moritz et al. (2015a). "Teaching Machines to Read and Comprehend". In: *CoRR* abs/1506.03340. arXiv: 1506.03340. URL: http://arxiv.org/abs/1506.03340.

Hochreiter, Sepp and Jürgen Schmidhuber (Nov. 1997). "Long Short-Term Memory". In: *Neural Comput.* 9.8, pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Hochreiter, Sepp et al. (2001). *Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-Term Dependencies*.

Jin, Rong and Alexander G. Hauptmann (2001). "Automatic Title Generation for Spoken Broadcast News". In: *Proceedings of the First International Conference on Human Language Technology Research*. URL: https://www.aclweb.org/anthology/H01-1011.

Kalchbrenner, Nal and Phil Blunsom (Oct. 2013). "Recurrent Continuous Translation Models". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, USA: Association for Computational Linguistics, pp. 1700–1709. URL: https://www.aclweb.org/anthology/D13-1176.

Klein, Guillaume et al. (July 2017b). "OpenNMT: Open-Source Toolkit for Neural Machine Translation". In: *Proceedings of ACL 2017, System Demonstrations*. Vancouver, Canada: Association for Computational Linguistics, pp. 67–72. URL: https://www.aclweb.org/anthology/P17-4012.

Klein, Guillaume et al. (2017a). "OpenNMT: Open-Source Toolkit for Neural Machine Translation". In: *CoRR* abs/1701.02810. arXiv: 1701.02810. URL: http://arxiv.org/abs/1701.02810.

Klein, Guillaume et al. (Mar. 2018). "OpenNMT: Neural Machine Translation Toolkit". In: *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers)*. Boston, MA: Association for Machine Translation in the Americas, pp. 177–184. URL: https://www.aclweb.org/anthology/W18-1817.

Kuratov, Yuri and Mikhail Arkhipov (2019). *Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language*. arXiv: 1905.07213 [cs.CL].

Lin, Chin-Yew (July 2004). "ROUGE: A Package for Automatic Evaluation of Summaries". In: *Text Summarization Branches Out*. Barcelona, Spain: Association for Computational Linguistics, pp. 74–81. URL: https://www.aclweb.org/anthology/W04-1013.

Liu, Yang and Mirella Lapata (2019). *Text Summarization with Pretrained Encoders*. arXiv: 1908.08345 [cs.CL].

Loper, Edward and Steven Bird (2002). "NLTK: The Natural Language Toolkit". In: *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*. ETMTNLP '02. Philadelphia, Pennsylvania: Association for Computational Linguistics, 63–70. DOI: 10.3115/1118108.1118117. URL: https://doi.org/10.3115/1118108.1118117.

Luong, Minh-Thang, Hieu Pham, and Christopher D. Manning (2015). *Effective Approaches to Attention-based Neural Machine Translation*. arXiv: 1508.04025 [cs.CL].

MacKay, David (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK New York: Cambridge University Press. ISBN: 978-0521642989.

Manning, Christopher and Richard Socher. *CS224n: NLP with Deep Learning*. URL: http://web.stanford.edu/class/cs224n/readings/cs224n-2019-notes05-LM_RNN.pdf.

Mihalcea, Rada and Paul Tarau (July 2004). "TextRank: Bringing Order into Text". In: *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain: Association for Computational Linguistics, pp. 404–411. URL: https://www.aclweb.org/anthology/W04-3252.

Mikolov, Tomas et al. (2013a). "Distributed Representations of Words and Phrases and their Compositionality". In: *CoRR* abs/1310.4546. arXiv: 1310.4546. URL: http://arxiv.org/abs/1310.4546.

Mikolov, Tomas et al. (2013b). *Efficient Estimation of Word Representations in Vector Space*. arXiv: 1301.3781 [cs.CL].

Mikolov, Tomas et al. (2017). *Advances in Pre-Training Distributed Word Representations*. arXiv: 1712.09405 [cs.CL].

Nallapati, Ramesh, Feifei Zhai, and Bowen Zhou (2016). *SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents*. arXiv: 1611.04230 [cs.CL].

Nallapati, Ramesh et al. (2016). *Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond*. arXiv: 1602.06023 [cs.CL].

Narayan, Shashi, Shay B. Cohen, and Mirella Lapata (2018). "Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, pp. 1797–1807. DOI: 10.18653/v1/D18-1206. URL: https://www.aclweb.org/anthology/D18-1206.

Nenkova, Ani (2005). "Automatic Text Summarization of Newswire: Lessons Learned from the Document Understanding Conference". In: *Proceedings of the 20th National Conference on Artificial Intelligence - Volume 3*. AAAI'05. Pittsburgh, Pennsylvania: AAAI Press, 1436–1441. ISBN: 157735236x.

Neubig, Graham (2017). *Neural Machine Translation and Sequence-to-sequence Models: A Tutorial*. arXiv: 1703.01619 [cs.CL].

Page, Lawrence et al. (1999). *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report 1999-66. Previous number = SIDL-WP-1999-0120. Stanford InfoLab. URL: http://ilpubs.stanford.edu:8090/422/.

Paul Over, James Yen (2001). "An introduction to duc-2001: Intrinsic evaluation of generic news text summarization systems." In:

— (2002). "An introduction to duc-2002: Intrinsic evaluation of generic news text summarization systems." In:

— (2003). "An introduction to duc-2003: Intrinsic evaluation of generic news text summarization systems." In:

Paulus, Romain, Caiming Xiong, and Richard Socher (2017). *A Deep Reinforced Model for Abstractive Summarization*. arXiv: 1705.04304 [cs.CL].

Peters, Matthew E. et al. (2018). *Deep contextualized word representations*. arXiv: 1802.05365 [cs.CL].

Radford, Alec (2018). "Improving Language Understanding by Generative Pre-Training". In:

Robertson, Stephen and Hugo Zaragoza (2009). "The Probabilistic Relevance Framework: BM25 and Beyond". In: *Foundations and Trends® in Information Retrieval* 3.4, pp. 333–389. ISSN: 1554-0669. DOI: 10.1561/1500000019. URL: http://dx.doi.org/10.1561/1500000019.

Rush, Alexander M., Sumit Chopra, and Jason Weston (2015). *A Neural Attention Model for Abstractive Sentence Summarization*. arXiv: 1509.00685 [cs.CL].

Sandhaus, Evan (2008). *The New York Times Annotated Corpus LDC2008T19*.

Secareanu, Alin (2017). "Football Events". In:

Segalovich, Ilya (Jan. 2003). "A Fast Morphological Algorithm with Unknown Word Guessing Induced by a Dictionary for a Web Search Engine." In: pp. 273–280.

Shannon, C. E. (1948). "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3, pp. 379–423. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Shavrina T., Shapovalova O (2017). "TO THE METHODOLOGY OF CORPUS CONSTRUCTION FOR MACHINE LEARNING: «TAIGA» SYNTAX TREE CORPUS AND PARSER". In: *"CORPORA2017", international conference , Saint-Petersbourg, 2017*.

SimilarWeb LTD (2019). *Top sites ranking for Sports in the world*. https://www.similarweb.com/top-websites/category/sports.

Sokolov, Andrej (2019). "Phrase-based attentional transformer for headline generation". In: *Computational Linguistics and Intellectual Technologies*.

Stepanov, Matvey (2019). "News headline generation using stems, lemmas and grammemes". In: *Computational Linguistics and Intellectual Technologies*.

Subedi, Nishan (2018). *FastText: Under the Hood*. URL: https://towardsdatascience.com/fasttext-under-the-hood-11efc57b2b3.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le (2014). *Sequence to Sequence Learning with Neural Networks*. arXiv: 1409.3215 [cs.CL].

Tan, Jiwei, Xiaojun Wan, and Jianguo Xiao (July 2017a). "Abstractive Document Summarization with a Graph-Based Attentional Neural Model". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume*

1: Long Papers). Vancouver, Canada: Association for Computational Linguistics, pp. 1171–1181. DOI: 10.18653/v1/P17-1108. URL: https://www.aclweb.org/anthology/P17-1108.

Tan, Jiwei, Xiaojun Wan, and Jianguo Xiao (2017b). "From Neural Sentence Summarization to Headline Generation: A Coarse-to-fine Approach". In: *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. IJCAI'17. Melbourne, Australia: AAAI Press, pp. 4109–4115. ISBN: 978-0-9992411-0-3. URL: http://dl.acm.org/citation.cfm?id=3171837.3171860.

— (2017c). "From Neural Sentence Summarization to Headline Generation: A Coarse-to-Fine Approach". In: *IJCAI*.

Uszkoreit, Jakob (2017). *Transformer: A Novel Neural Network Architecture for Language Understanding*. URL: https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html.

Vaswani, Ashish et al. (2017). *Attention Is All You Need*. arXiv: 1706.03762 [cs.CL].

Venugopalan, Subhashini et al. (2015). *Sequence to Sequence – Video to Text*. arXiv: 1505.00487 [cs.CV].

Vinyals, Oriol, Meire Fortunato, and Navdeep Jaitly (2015). "Pointer Networks". In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes et al. Curran Associates, Inc., pp. 2692–2700. URL: http://papers.nips.cc/paper/5866-pointer-networks.pdf.

Wikipedia contributors (2019). *List of most-watched television broadcasts — Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=List_of_most-watched_television_broadcasts&oldid=927501490. [Online; accessed 24-November-2019].

Wong, Kam-Fai, Mingli Wu, and Wenjie Li (Aug. 2008). "Extractive Summarization Using Supervised and Semi-Supervised Learning". In: *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*. Manchester, UK: Coling 2008 Organizing Committee, pp. 985–992. URL: https://www.aclweb.org/anthology/C08-1124.

Zhang, Xiang, Junbo Zhao, and Yann LeCun (2015). *Character-level Convolutional Networks for Text Classification*. arXiv: 1509.01626 [cs.LG].