Text Summarization of Czech News Articles Using Named Entities

Petr Marek, Štěpán Müller, Jakub Konrád, Petr Lorenc, Jan Pichl, Jan Šedivý

Faculty of Electrical Engineering, CTU in Prague, Prague, Czech Republic

Abstract

The foundation for the research of summarization in the Czech language was laid by the work of Straka et al. (2018). They published the SumeCzech, a large Czech news-based summarization dataset, and proposed several baseline approaches. However, it is clear from the achieved results that there is a large space for improvement.

In our work, we focus on the impact of named entities on the summarization of Czech news articles. First, we annotate SumeCzech with named entities. We propose a new metric ROUGE$_{\text{NE}}$ that measures the overlap of named entities between the true and generated summaries, and we show that it is still challenging for summarization systems to reach a high score in it.

We propose an extractive summarization approach *Named Entity Density* that selects a sentence with the highest ratio between a number of entities and the length of the sentence as the summary of the article. The experiments show that the proposed approach reached results close to the solid baseline in the domain of news articles selecting the first sentence. Moreover, we demonstrate that the selected sentence reflects the style of reports concisely identifying *to whom*, *when*, *where*, and *what* happened. We propose that such a summary is beneficial in combination with the first sentence of an article in voice applications presenting news articles.

We propose two abstractive summarization approaches based on Seq2Seq architecture. The first approach uses the tokens of the article. The second approach has access to the named entity annotations. The experiments show that both approaches exceed state-of-the-art results previously reported by Straka et al. (2018), with the latter achieving slightly better results on SumeCzech's out-of-domain testing set.

1. Introduction

Automatic text summarization is an important task of natural language understanding. The goal is to describe a text accurately, be it a news article, a web page, or a paragraph of a book, using shorter text. The shorter text can be in the form of a paragraph, sentence, or even a few words. Automatic text summarization is a challenging problem for automatic systems because they have to excel in multiple areas at once. They have to understand the meaning of the original text, understand which passages are important and which can be excluded, and generate meaningful and grammatically correct summarizations.

In this work, we focus on the summarization of Czech news articles by a one-sentence summary. We can also describe this task as the automatic creation of a headline for a given text. We use the SumeCzech dataset (Straka et al., 2018) for our experiments.

Additionally, we explore the influence of the named entities on text summarization. We use SpaCy's named entity recognition (NER) model, trained on a CoNLL-based extended CNEC 2.0 dataset (Ševčíková et al., 2014), to label SumeCzech with named entities to create additional features. We publish the annotations to promote the replication of results and to enable further research.

We use the annotations as a foundation for our newly proposed *Named Entity Density*. The method selects the sentence with the highest ratio of the number of entities to the sentence length as the summary of the article. We show that our proposed method achieves nearly as good results in the automatic evaluation as the hard-to-beat baseline in the news domain that selects the first sentence. Nenkova (2005) shows that the baseline selecting the first sentence is a strong baseline because authors tend to summarize the main points of an article in the first sentence, especially in the news domain. We also show that sentences selected by *Named Entity Density* possess a high information value mentioning *to whom*, *where*, *when*, and *what* happened. This structure resembles the style of reports that concisely identifies and examines issues, events, or findings that have happened. We propose that such a summary is useful in voice applications presenting news. Voice application can present the summary formed out of the first sentence of a news article first and continue with the sentence selected by *Named Entity Density* if a user requests additional information.

We also propose two abstractive methods that can construct a novel sentence as a summary. They are based on the Seq2Seq architecture, initially used for machine translation. The first method uses the text of the article only. The second method uses additional annotations created by the name entity recognition system as input features. Our experiments show that both models achieved state-of-the-art results in SumeCzech's task to summarize the headline from the text of the article. We also show that the named entities added as an additional input feature improve the ability of the model to generalize to the out-of-domain data.

Finally, we propose a new metric Rogue$_{NE}$, which measures the overlap of named entities in the target and generated summaries. Poor results of the experiments in Rogue$_{NE}$ show that summarization of entities still poses many challenges, and this task has not been solved yet.

2. Related Work

Allahyari et al. (2017) provide a brief survey of text summarization. In general, we divide text summarization algorithms into two categories, *extractive* and *abstractive*.

2.1. Extractive Summarization

Extractive summarization algorithms choose pieces from the original text, usually sentences, and combine them to form a summary. From a high-level perspective, most extractive summarizers follow the same two steps: First, score all sentences. Then, pick N sentences with the highest score. The main difference between individual extractive methods is how they score sentences. The advantage of extractive methods is that no matter how simple the method is, it always produces syntactically correct sentences, even though they may not be useful summaries. On the other hand, there is a disadvantage too. Extractive summarizers are limited in what they can predict by the sentences of the source text. Thus, more elaborate summaries are out of their reach.

Mihalcea and Tarau (2004) introduce a Textrank algorithm, a graph-based ranking model. It creates a graph of sentences based on their overlap. It chooses the most important sentences according to the created graph. Pal and Saha (2014) propose a summarization algorithm that derives the relevance of the sentences within the text using the Simplified Lesk algorithm and the WordNet online database. Kågebäck et al. (2014) propose using continuous vector representations for semantically aware representations of sentences for summarization. Zhang et al. (2016) develop convolutional neural networks that learn sentence features and perform sentence ranking. The latest results are achieved by Liu (2019). They apply the BERT model (Devlin et al., 2018) to extractive summarization.

Especially relevant works for our research are those working with named entities. Nobata et al. (2002) introduce named entity tagging and pattern discovery to a summarization system based on a sentence extraction technique. Hassel (2003) integrates a Named Entity tagger into the SweSum summarizer for Swedish newspaper texts. Filatova and Hatzivassiloglou (2004) propose a summarization technique using a set of features based on low-level, atomic events that describe the relationships between important actors in a document or in a set of documents. The extraction of atomic events relies on a noun phrase and named entity recognition (Hatzivassiloglou and Filatova, 2003). Jabeen et al. (2013) apply named entity recognition for summarization of tweets. Schulze and Neves (2016) present EntityRank, a multidocument

graph-based summarization algorithm that is solely based on named entities. They apply it to texts from the medical domain successfully. Khademi and Fakhredanesh (2020) propose an unsupervised method for summarizing Persian texts that use a named entity recognition system. Their method consists of three phases: training a supervised NER model, recognizing named entities in the text, and generating a summary.

2.2. Abstractive Summarization

Abstractive summarizers generate summarizations consisting of novel sentences that were not part of the original text. Abstractive summarization algorithms are usually more complex because they have to understand the input text, find the most relevant passages, and generate syntactically correct sentences as summarization. Such a task is nearly impossible for hand-written rules. However, the recent advance of machine learning and, in particular, neural networks makes abstractive summarization possible. Moreover, neural networks represent the current state-of-the-art in abstractive summarization.

Nallapati et al. (2016) models abstractive text summarization using Attentional Encoder-Decoder Recurrent Neural Networks. They propose several novel models that address critical problems in summarization that are not adequately modeled by the basic architecture, such as modeling keywords, capturing the hierarchy of sentence-to-word structure, and emitting rare or unseen words during the training time. Liu et al. (2017) propose an adversarial process for abstractive text summarization. Yao et al. (2018) propose a recurrent neural network-based Seq2Seq attentional model with a dual encoder including the primary and the secondary encoders. Song et al. (2019) propose an LSTM-CNN based approach that can construct new sentences by exploring more fine-grained fragments than sentences, namely, semantic phrases. The proposed approach is composed of two main stages. The first stage extracts phrases from source sentences. The second stage generates text summaries using deep learning. Liu and Lapata (2019) apply BERT in text summarization and propose a general framework for both extractive and abstractive models. For abstractive summarization, they propose a new fine-tuning schedule that adopts different optimizers for the encoder and the decoder as a means of alleviating the mismatch between the two as the former is pretrained while the latter is not.

To the best of our knowledge, there is no work exploring the influence of named entities on the extractive summarization techniques, let alone in the Czech language.

3. Dataset

We use SumeCzech for experiments. SumeCzech is Czech news-based summarization dataset created by Straka et al. (2018). It contains more than a million documents, consisting of a headline, several sentences long abstract, and a full text. The

Website	Number	Percentage
ceskenoviny.cz	4,854	0.5%
denik.cz	157,581	15.7%
idnes.cz	463,192	46.2%
lidovky.cz	136,899	13.7%
novinky.cz	239,067	23.9%
Total	1,001,593	100.0%

Table 1. The number of documents in SumeCzech from individual news websites

dataset was collected from various Czech news websites. We show the distribution of the websites in Table 1.

SumeCzech is split into four parts. Three of them are the train, development, and test sets. Additionally, to simulate a real-life situation where a model is trained on data from one domain, and used on real data from other domains, Straka et al. (2018) created an out-of-domain (OOD) test set. OOD test set evaluates how models cope with news articles from domain never seen during training. They clustered the whole dataset into 25 clusters using K-Means on abstracts of the articles and selected one cluster as the OOD test set. The OOD test set contains approximately 4.5% of all articles. The OOD testing set seems to contain news articles about concerts and festivals. The remaining articles were divided into train, development, and test sets in 86.5 : 4.5 : 4.5 ratio.

3.1. Named Entity Annotations

We train a model for named entity recognition in the Czech language to annotate SumeCzech by named entities. We selected the CoNLL-based extended CNEC 2.0 (Konkol et al., 2014) as the training dataset, as it is the largest and most up-to-date Czech named entity recognition dataset. The advantage is that the dataset contains no nested entities, making the outputs easier to use for summarizers.

We selected SpaCy's NER model[1] (Honnibal et al., 2020) because previous experiments by Müller (2020a) showed that SpaCy's NER model offers a good trade-off between performance, speed, and memory requirements. Speed and memory requirements might seem unimportant for our experiments because we can precompute the annotations. However, for the sake of practical usage, in which the labels have to be created as soon as possible once a new document for summarization arrives, we decided to take those properties into account too. The SpaCy's NER model achieved a 78.45 F-Score on the testing set of CoNLL-based extended CNEC 2.0. For compar-

---

[1]https://spacy.io/api/entityrecognizer

Entity Type	Train	Dev	Test	Test OOD
Numbers in addresses	116,990	5,052	5,129	1,827
Geographical names	5,271,938	282,440	285,307	212,637
Institutions	4,488,357	222,524	234,147	250,555
Media names	534,340	24,379	27,966	22,360
Artifact names	2,367,532	118,938	108,811	196,009
Personal names	7,991,790	406,938	395,867	646,556
Time expressions	1,684,152	87,096	86,866	121,357
Total	22,455,099	1,147,367	1,144,093	1,451,301

Table 2. Number of named entities in texts of SumeCzech's articles

Entity Type	Train	Dev	Test	Test OOD
Numbers in addresses	331	18	12	26
Geographical names	285,148	15,903	14,697	13,502
Institutions	161,809	7,578	8,472	12,806
Media names	9,088	371	420	718
Artifact names	62,124	3,344	2,837	7,748
Personal names	302,276	15,117	15,856	31,266
Time expressions	14,400	760	838	1,127
Total	835,176	43,091	43,132	67,193

Table 3. Number of named entities in headlines of SumeCzech's articles

Entity Type	Train	Dev	Test	Test OOD
Numbers in addresses	1,686	105	85	83
Geographical names	773,901	41,759	38,903	33,001
Institutions	601,129	28,380	33,119	52,938
Media names	77,591	3,744	4,320	3,946
Artifact names	159,122	7,550	7,174	25,204
Personal names	747,686	36,783	37,712	65,950
Time expressions	132,276	7,214	7,272	23,544
Total	2,493,391	125,535	128,585	204,666

Table 4. Number of named entities in abstracts of SumeCzech's articles

sion, current state-of-the-art result on this dataset is 86.39 F-Score (Straková et al., 2019; Müller, 2020b).

We applied the trained SpaCy's NER model to the text of SumeCzech's articles. The result was annotations in IOB2 format, one label for each word token. The NER found approximately 26M named entities in texts, 1M in headlines, and 3M in abstracts. (We do not use abstracts in our experiments. We present the numbers of named entities in the abstracts for completeness only.) We show the detailed statistics in Table 2, Table 3, and Table 4. We also counted the number of headlines without any named entity. We show the statistic in Table 5. We published the annotations[2] to promote replication of results and to enable further research (Marek and Müller, 2021).

Split	Percentage
Train	36.1%
Dev	35.4%
Test	35.7%
Test OOD	14.1%

Table 5. Percentage of headlines containing no named entity.

4. Metrics

We used the ROUGE$_{RAW}$ metric for evaluation. ROUGE$_{RAW}$ was proposed by Straka et al. (2018) as a language-agnostic variant of ROGUE (Lin, 2004). The original ROGUE metric automatically determines the quality of the generated summary by comparing it to a reference summary created by humans. There are two variants. ROUGE-N measures the overlap of N-grams between the generated and reference summaries. ROUGE-L looks at the longest common subsequence between the reference and the generated summaries. ROGUE calculates recall and is English-specific. It employs English stemmer, stop words, and synonyms.

ROUGE$_{RAW}$ does not need any stemmer, stop words, or synonyms, which makes it language independent. It measures recall, precision and F-score. It also has two variants ROUGE$_{RAW}$-N and ROUGE$_{RAW}$-L corresponding to the variants of the original ROGUE metric. We selected ROUGE$_{RAW}$-1, ROUGE$_{RAW}$-2, and ROUGE$_{RAW}$-L to evaluate approaches that we propose.

---

[2]http://hdl.handle.net/11234/1-3505

4.1. ROUGE$_{NE}$

Since we focus on the role of named entities in summarization, we propose a novel metric ROUGE$_{NE}$. ROUGE$_{NE}$ measures the overlap of named entities between the reference and the generated summaries. This metric evaluates the ability of the model to transfer named entities to the summary.

Formally, let us denote the tokens of a true summary X:

$$X = \{x_0, x_1, x_2, \ldots, x_n\},$$

where x_i are individual tokens of the summary. Let us denote the generated summary Y in a similar fashion:

$$Y = \{y_0, y_1, y_2, \ldots, y_m\}.$$

Next, we apply the named entity recognition algorithm on X and Y. The result is entity annotations xe_i and ye_i for all tokens x_i and y_i:

$$\{xe_0, xe_1, xe_2, \ldots, xe_n\},$$

$$\{ye_0, ye_1, ye_2, \ldots, ye_m\}.$$

The annotations can be divided into a set of annotations E_{NE}, that mark entities, and annotations $E_{\neg NE}$, that do not mark entities. In the analogy of IOB format, the former are I and B annotations, and the latter are O annotations. For the calculation of ROUGE$_{NE}$ we select only tokens that are marked as entities. Formally, we select only the tokens of summaries X and Y, for which its entity label is an element of E_{NE}. The results are the X_e and Y_e:

$$X_e = \{x_i\} \text{ for } i = 0 \ldots n \text{ if } xe_i \in E_{NE},$$

$$Y_e = \{y_i\} \text{ for } i = 0 \ldots m \text{ if } ye_i \in E_{NE}.$$

Next, we calculate the ROGUE precision and recall scores using the tokens of X_e and Y_e as follows:

$$precision = \frac{|X_e \cap Y_e|}{|X_e|},$$

$$recall = \frac{|X_e \cap Y_e|}{|Y_e|},$$

where $|X_e|$ and $|Y_e|$ denote the sizes of X_e and Y_e. $|X_e \cap Y_e|$ denotes the number of overlapping tokens between X_e and Y_e. The resulting values are the precision and recall of ROUGE$_{NE}$. We define the metrics to be equal to zero for summaries without any named entity.

5. Methods

The task we study is to create a one-sentence summary from the text of the article. The one-sentence summarization can be seen as the task to create a headline of the article. We use five baselines introduced by Straka et al. (2018). Moreover, we propose one extractive method – *Named Entity Density* and two abstractive approaches, *Seq2Seq* and *Seq2Seq–NER*, for text summarization.

5.1. Baselines

We adopt the methods proposed by Straka et al. (2018) as a baseline. They propose four extractive and one abstractive methods for SumeCzech's task to create a headline out of the text of the article:

- *First*: unsupervised extractive method. It returns the first sentence of the article.
- *Random*: unsupervised extractive method. It returns a random sentence from the article.
- *TextRank*: unsupervised extractive method. It selects the most important sentence of the article using the TextRank (Mihalcea and Tarau, 2004) algorithm.
- *clf-rf*: supervised extractive method. It selects the sentence that receives the highest score produced by the Random forest classifier. The classifier performs classification using vector representation of sentences. The vector representation consists of the sum of TF-IDF for each word normalized by the sentence length, length of the sentence, cohesion (distance from other sentences), the count of capitalized words in the sentence, the count of tokens that consist of digits, and the count of non-essential words that suggests the sentence relates to some other sentence.
- *t2t*: supervised abstractive method. It uses a neural machine translation model proposed by Vaswani et al. (2017) to generate a summary consisting of a novel sentence.

5.2. Named Entity Density

Named Entity Density is our proposed unsupervised extractive method. It calculates the named entity density score for each sentence and selects the sentence with the highest score. The score is calculated in two steps. First, we apply a named entity recognition algorithm to all sentences. Next, we calculate the named entity density as a ratio of the number of tokens marked as a named entity to the total number of tokens in the sentence.

Formally, let us denote the article A, for which we want to create a summary, as a set of sentences $s_0 \ldots s_n$:

$$A = \{s_0, s_1, s_2, \ldots, s_n\}.$$

Each sentence s_i contains word tokens $x_0 \ldots x_m$:

$$s_i = \{x_0, x_1, x_2, \ldots, x_m\}.$$

We apply the named entity recognition algorithm NER on each sentence s_i of the article A:

$$\text{NER}(s_i) = \{e_0, e_1, e_2, \ldots, e_m\},$$

that produces NER labels $e_0 \ldots e_m$ for each token $x_0 \ldots x_m$ of the sentence s_i. We can divide the NER tokens into two sets: E_{NE} and $E_{\neg NE}$. Each NER token belongs into exactly one of those two sets. E_{NE} contains all NER tokens representing some entity type. $E_{\neg NE}$ contains all NER tokens that do not represent any entity type. In the analogy of IOB format, E_{NE} contains I and B tokens, and $E_{\neg NE}$ contains O tokens. Next, we calculate the named entity density NED for each sentence $s_0 \ldots s_n$ of the article A. The NED is defined as:

$$\text{NED}(s_i) = \frac{|E_{NE}|}{|s_i|},$$

where $|E_{NE}|$ denotes the number of tokens in the sentence s_i which NER algorithm marked as named entities and $|s_i|$ is the number of all tokens $x_0 \ldots x_m$ forming the sentence s_i. We select the sentence s_i with the highest NED score as a summary of article A.

The intuition of the *Named Entity Density* is that the sentence with the high NED score mentions the highest number of entities within the smallest text fragment. Such a sentence corresponds to the form of a report that is structured around concisely identifying and examining issues, events, or findings that have happened.

5.3. Seq2Seq

Seq2Seq is a supervised abstractive method that uses a Seq2Seq model with global attention. Formally, a Seq2Seq neural network models the conditional probability $p(y|x)$ of translating a source text $x = \{x_0, x_1, x_2, \ldots, x_n\}$ into a target text $y = \{y_0, y_1, y_2, \ldots, y_m\}$ (Luong et al., 2015). The source text x is an article, y is a summary, and $m < n$ in our case. The Seq2Seq neural network consists of an encoder and a decoder. The encoder creates a fixed-length vector representation r of the source text x:

$$r = \text{ENC}(x).$$

The encoder is usually a recurrent neural network with hidden states h_s^{ENC}:

$$h_s^{ENC} = f^{ENC}(h_{s-1}^{ENC}, x_s),$$

and the output of the encoder is its last hidden state:

$$\text{ENC}(x) = h_n^{ENC}$$

The f^{ENC} can be a vanila RNN (Rumelhart et al., 1985), LSTM (Hochreiter and Schmidhuber, 1997), or GRU (Cho et al., 2014) unit.

The decoder takes r as an input and generates a target text, one token at a time:

$$\log p(y|x) = \sum_{j=1}^{m} \log p(y_j|y_{<j}, r).$$

We can also represent the probability of generating a target word y_j as:

$$p(y_j|y_{<j}, r) = softmax(g(h_j^{DEC})),$$

where g is a transformation function that generates vocabulary sized vector. The h_j^{DEC} is the output of recurrent neural network unit:

$$h_j^{DEC} = f^{DEC}(h_{j-1}^{DEC}, y_{j-1}).$$

Function f^{DEC} can be a vanilla RNN, LSTM, or GRU unit like in the case of the encoder.

We add a global attention mechanism to the Seq2Seq neural network. The attention allows the network to focus on parts of the source text selectively during the target text generation. We illustrate the global attention mechanism in Figure 1.
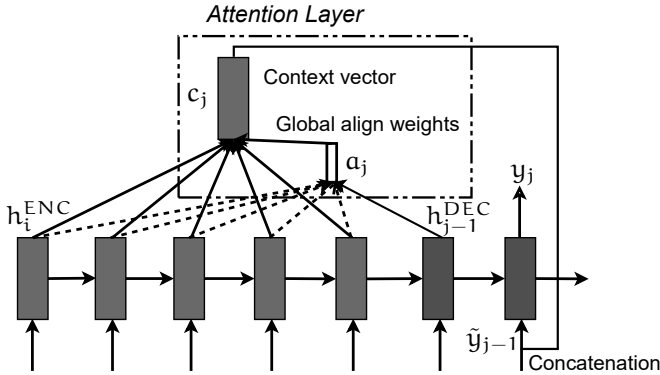

Figure 1. Seq2Seq with global attention mechanism. The figure is inspired by Luong et al. (2015).

The idea is to concatenate a source-side context vector c_j with the input vector of the encoder y_{j-1}:

$$\tilde{y}_{j-1} = [c_j, y_{j-1}].$$

The vector $\tilde{y}_{j-1}$ is fed into f^{DEC}:

$$h_j^{DEC} = f^{DEC}(h_{j-1}^{DEC}, \tilde{y}_{j-1}).$$

The context vector c_j is computed as a weighted average over all vectors of hidden states of the encoder h_s^{ENC}:

$$c_j = \sum_{i=0}^{n} a_{j_i} \cdot h_i^{ENC},$$

where a_{j_i} is the ith element of the weight vector a_j. The a_j is calculated by the softmax function comparing each hidden state of the encoder h_s^{ENC} with the current hidden state of the decoder h_j^{ENC}:

$$a_{j_i} = \frac{exp(score(h_{j-1}^{DEC}, h_i^{ENC}))}{\sum_{s'} exp(score(h_{j-1}^{DEC}, h_{s'}^{ENC}))}.$$

There are multiple definitions of the $score$ function. We selected the $general\ score$ function, defined as:

$$score(h_t, h_s) = h_t^\top W h_s.$$

5.4. Seq2Seq–NER

Seq2Seq–NER is a supervised abstractive method that uses a Seq2Seq model with global attention and adds the NER feature encoded as one-hot encoded vector appended to input embedding vector. Formally, the Seq2Seq neural network models the conditional probability $p(y|x_{NER})$, where y is a target text $y = \{y_0, y_1, y_2, ..., y_m\}$ and x_{NER} is a source sequence. The source sequence is a concatenation of a vector representation of the source token x_i and one-hot vector representation of entity type e_i:

$$x_{NER} = \{[x_0, e_0], [x_1, e_1], [x_2, e_2], ..., [x_n, e_n]\}.$$

The rest of the model works in a similar fashion as a Seq2Seq model, that we described in the subsection 5.3. To summarize, the difference between *Seq2Seq* and *Seq2Seq–NER* models is that the latter has access to the named entity labels of the source words produced by the NER algorithm.

6. Implementation Details

We implemented the baseline methods *first* and *random* proposed by Straka et al. (2018). We replicated results using ROUGE$_{RAW}$-1, ROUGE$_{RAW}$-2, and ROUGE$_{RAW}$-L metrics, and additionally evaluated ROUGE$_{NE}$ metric.

For the proposed supervised methods, we used a modified implementation of a Seq2Seq model with global attention from the official PyTorch tutorial (Weidman, 2019). The hidden sizes of the encoder and decoder were set to 256. The size of our vocabulary was 25,000. We used 300-dimensional fastText for embedding words. We used dropout 0.1 on the outputs of both RNNs. We trained our models until

the validation loss started increasing, and we selected the weights having the lowest validation loss for evaluation.

We encountered problems with attention. The implementation in the tutorial used attention similar to *concat global attention*. The tutorial suggested to copy hidden state of the decoder for each hidden state of the encoder, then concatenate the hidden states with the encoder outputs, and pass them to a linear layer to calculate energy. The reason was to utilize parallel nature of GPU. We used batch size 16. Therefore, the concatenated tensor would have 147,890,688 floating numbers in the case of the longest sentence. To train our model with attention without running out of memory on our GPU, we had to simplify the way attention was calculated. We used an approach similar to *general global attention*. We used an affine transformation on the hidden state of the decoder to transform it into a 64-dimensional vector to calculate energy. We also affinely transformed the encoder's hidden states vectors of the same dimension by a linear layer. The transformed decoder hidden state was then used as a multiplier and broadcast over all encoder hidden states, making the calculation of energy much more memory efficient because the hidden state of the decoder did not have to be copied.

A limited vocabulary of the model led to many unknown words in the titles, and the model that used word tokenization learned to predict them. We had to forbid the model from predicting unknown tokens during evaluation to get meaningful titles.

For the *Seq2Seq–NER* model, we encoded the entities using the IOB2 format. The format has one common *outside* tag, and *beginning* and *inside* tags for each entity type. We encoded NER features into a one-hot vector for each word. The vector has 17 dimensions. Fourteen dimensions are reserved for the beginning and inside tags for each of the seven entity types our NER distinguishes. One dimension represents the outside of the entity tag. One represents padding, and one represents both the start and end of sequence symbols. We concatenated the NER feature vector with the embedding vector entering the encoder of the Seq2Seq.

7. Results of Automatic Evaluation and Discussion

We show the results of the evaluation in Table 6. First, we replicated the results of *First* and *Random* baselines reported in Straka et al. (2018). Our results were on par with the reported Precision, Recall, and F-Score of Rogue$_{Raw}$-1, Rogue$_{Raw}$-2, Rogue$_{Raw}$-L. We additionally evaluated our proposed metric Rogue$_{NE}$ for comparison with other methods.

Next, we evaluated the proposed extractive method *Named Entity Density* (NE Density). Results of *Named Entity Density* compared to the *First*, a solid summarization baseline, especially in the news articles domain, are encouraging. *Named Entity Density* achieves only slightly worse results. Moreover, the achieved results are consistent between the test and the OOD test set. Additionally, as we will show in section 8, *Named Entity Density* produces summaries resembling the style of informationally concise reports.

We evaluated the *Seq2Seq* and *Seq2Seq–NER* on the Test set to compare those methods with the approaches proposed by Straka et al. (2018). We can see that our proposed *Seq2Seq* and *Seq2Seq–NER* methods achieve better results on average by 80% relatively in Precision and F-score compared to the best methods proposed by Straka et al. (2018). Only *Textrank* and *First* achieve better results in Recall. The *Seq2Seq–NER* achieved slightly better results than *Seq2Seq*, which proves NER labels' usefulness for summarization. Although, it seems from the results of Rogue$_{NE}$ that the better score is not caused by the improved performance of using entities in the summaries.

We evaluated the *Seq2Seq* and *Seq2Seq–NER* on the OOD test set to learn how models cope with news articles from a domain never seen during training. The results are encouraging. Even though they show a drop in absolute values of metrics between Test and OOD test sets, the trend is the same. *Seq2Seq* and *Seq2Seq–NER* methods achieve the best results of all compared methods in Precision and F-score, and *Seq2Seq–NER* has slightly better results than *Seq2Seq*.

Finally, we take a look at the results in Rogue$_{NE}$. We do not have results for *Textrank* and *Tensor2Tensor* because they were not reported in the work of Straka et al. (2018) and we did not implement the methods ourselves. However, it is clear from the rest of the results that even the recent state-of-the-art methods are struggling with the named entities in the summarization.

Dataset	Method	Rogue$_{Raw}$-1			Rogue$_{Raw}$-2			Rogue$_{Raw}$-L			Rogue$_{NE}$		
		P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
Test	First	7.8	14.6	9.4	1.1	*2.3*	1.5	6.7	12.6	8.1	2.4	2.7	2.4
	Random	6.2	11.0	7.3	0.5	0.9	0.6	5.4	9.5	6.3	1.8	2.1	1.8
	Textrank	6.0	*16.5*	8.3	0.8	0.6	0.7	5.0	*13.8*	6.9	-	-	-
	Tensor2Tensor	8.8	7.0	7.5	0.8	0.6	0.7	8.1	6.5	7.0	-	-	-
	NE Density	6.6	10.7	7.3	0.8	1.4	0.9	5.9	9.4	6.4	1.5	2.2	1.6
	Seq2Seq	16.1	14.1	14.6	*2.5*	2.1	*2.2*	14.6	12.8	13.2	*5.3*	*6.5*	*5.6*
	Seq2Seq–NER	*16.2*	14.1	*14.7*	*2.5*	2.1	*2.2*	*14.7*	12.8	*13.3*	4.7	6.0	5.0
OOD	First	7.0	14.7	8.7	1.4	*2.9*	1.7	6.1	*12.8*	7.6	*1.4*	1.7	*1.4*
	Random	5.5	10.9	6.6	0.7	1.4	0.8	4.8	9.5	5.8	0.9	1.3	1.0
	Textrank	5.8	*16.9*	8.1	1.1	3.4	1.5	5.0	14.5	6.9	-	-	-
	Tensor2Tensor	6.3	5.1	5.5	0.5	0.4	0.4	5.9	4.8	4.8	-	-	-
	NE Density	6.3	11.4	7.1	1.3	2.3	1.4	5.7	10.2	6.3	1.0	*1.9*	1.1
	Seq2Seq	13.1	11.8	12.0	*2.0*	1.7	*1.8*	12.1	11.0	11.2	1.0	1.0	1.0
	Seq2Seq–NER	*13.7*	11.9	*12.4*	*2.0*	1.7	*1.8*	*12.6*	11.1	*11.4*	0.9	0.9	0.9

Table 6. Results of automatic evaluation

8. Examples

We choose a few representative examples from the test and OOD test sets to show how different methods summarize. We also provide English translation for convenience. Only very simple automatic post-processing was done on the output of the proposed *Seq2Seq* and *Seq2Seq–NER* models. We filtered the start of the sentence and end of the sentence symbols, removed spaces before punctuation, stripped the text of any starting or ending space, and capitalized the first letter.

First, we present examples of summarization created by *Named Entity Density* in Table 7. We do not divide the examples into test and OOD test sets because the generated summaries of both sets achieve comparative quality thanks to the fact that *Named Entity Density* is an unsupervised method.

We can see that the selected sentences contain named entities. Those sentences are comprised of factual information. The sentences are always grammatically correct thanks to the fact that *Named Entity Density* is an extractive approach. Even though the summaries created by *First* can contain more entities in general, the summaries created by *Named Entity Density* have a higher density of entities. We can see that the summaries created by *Named Entity Density* revolve around *to whom*, *when*, *where*, and *what* happened. It closely resembles the style of reports that concisely identify and examine issues, events, or findings that have happened.

Notice also that the sentences selected by *Named Entity Density* are not the first sentences of the articles. We measured that the sentence selected as a summary by *Named Entity Density* differs from the sentence selected by *First* in 93% of SumeCzech's articles. Thus, we can use the summaries created by *Named Entity Density* as an alternative version or reformulation of summaries created by method selecting the *First* sentence of an article. This property is highly praised by voice applications like Alquist (Pichl et al., 2020) or Emora (Finch et al., 2020). Voice applications present news articles in a summary because users quickly lose focus as news articles are not intended to be read by synthetic voices. Initially, the voice application can present the first sentence of the article. Additionally, if the user requests to learn more, it can present the summary produced by *Named Entity Density*.

We show the results of *Seq2Seq* and *Seq2Seq–NER* models for test and OOD test sets separately in Table 8 and Table 9. Both models generate novel sentences and incorporate entities into the generated summarizations successfully. We can see that despite the promising results of the automatic evaluation, a part of the outputs are not grammatically correct and contain repeated words.

9. Conclusion

This work explored the summarization of Czech news articles and influence of named entity labels for this task. We selected the SumeCzech dataset for our experiments. SumeCzech is over one million articles large dataset collected by Straka et al.

Method	Headline
Gold	Maloobchod v srpnu výrazně rostl *Retail trade grew significantly in August*
First	Po očištění o sezónní a kalendářní vlivy rostl maloobchod meziročně o 4,2 procenta. *After adjusting for seasonal and calendar effects, retail trade grew by 4.2 percent year on year.*
NED	Podle Eurostatu vzrostly meziročně kalendářně očištěné maloobchodní tržby v celé Evropské unii o 2,2 procenta. *According to Eurostat, calendar-adjusted retail sales rose by 2.2 percent year on year across the European Union.*
Gold	Snoubenci zestárli, přibývá levnějších obřadů bez svatebčanů *The couple is getting old, there are more and more cheaper ceremonies without wedding guests*
First	Stoupá počet sňatků bez svatebčanů, ve všední den, jen za přítomnosti svědků. *The number of marriages without wedding guests is increasing, on weekdays, only in the presence of witnesses.*
NED	Centrum metropole bude stále patřit k nejžádanějším místům pro oddávání, potvrdila Právu mluvčí Prahy 1 Veronika Blažková. *The center of the metropolis will still be one of the most sought-after places for wedding, "Veronika Blažková, spokeswoman for Prague 1, confirmed to Právo.*
Gold	Vranovskou přehradu znovu znečistila ropa, unikala ze sudů na dně *The Vranov dam was again polluted by oil, escaping from barrels at the bottom*
First	Likvidace probíhá za odborné spolupráce pracovníků povodí Moravy a odboru životního prostředí. *The liquidation takes place with the professional cooperation of the employees of the Moravia River Basin and the Department of the Environment.*
NED	Starosta Vranova nad Dyjí se o ropě dozvěděl z tisku, což jej rozlítilo. *The mayor of Vranov nad Dyjí learned about the oil from the press, which angered him.*
Gold	Z Fondové bude Reaganova žena, doplní ji Oprah a Poslední skotský král *The Fond will be Reagan's wife, complemented by Oprah and the Last King of Scotland*
First	Ve snímku s názvem The Butler (Majordomus) o správci v Bílém domě pracujícím pro několik amerických prezidentů by se v hlavní roli mohl podle časopsisu (the word *časopsisu* is misspelled in the dataset) Variety objevit americký herec Forest Whitaker. *According to Variety magazine, American actor Forest Whitaker could star in the film The Butler about a White House caretaker working for several US presidents.*
NED	Amerického prezidenta Richarda Nixona si zřejmě zahraje John Cusak. *US President Richard Nixon is likely to be played by John Cusack.*
Gold	Zlatého ledňáčka na festivalu Finále Plzeň získal snímek Jako nikdy *Movie Jako nikdy won Golden Kingfisher at the Finale Pilsen festival*
First	Letošní ročník festivalu Finále byl výjimečný tím, že poprvé soutěžily kromě českých také slovenské snímky. *This year's Finale festival was exceptional in that, for the first time, in addition to Czech, Slovak films also competed*
NED	Letošní ročník festivalu Finále Plzeň navštívilo od 27. dubna do 3. května 10 853 diváků. *This year's Finale Plzeň festival was visited by 10,853 spectators from April 27 to May 3.*

Table 7. Examples of summarizations created by Named Entity Density

Method	Headline
Gold	Nejznámější Albánec může o stavbě mešity přemýšlet ve vězení
	The most famous Albanian can think about building a mosque in prison
Seq2Seq	Soud potrestal za únos s lidmi
	The court punished for kidnapping with people
Seq2Seq–NER	Soud potvrdil tresty za pašování drog
	The court upheld the penalties for drug smuggling
Gold	Kriminalisté dopadli násilníka, který v lednu zneužil školáky z Orlové
	Criminal investigators caught a rapist who abused schoolchildren from Orlová in January
Seq2Seq	Policie hledá muže, který v Ostravě znásilnil děti
	Police are looking for a man who raped children in Ostrava
Seq2Seq–NER	Policie hledá muže, který se v Ostravě, který se na něj
	Police are looking for a man in Ostrava who at him
Gold	Do Valtického Podzemí za divadlem místo vína
	To the Valtice Underground for the theater instead of wine
Seq2Seq	Divadlo se v Brně otevře v Brně
	The theater in Brno will open in Brno
Seq2Seq–NER	V Brně otevřeli novou sezonu, divadlo se otevře návštěvníkům
	They have opened a new season in Brno, the theater will be open to visitors

Table 8. Examples of summarizations from the Test set

(2018) from Czech news websites. We annotated SumeCzech by named entities by the SpaCy's NER. We published the annotations to promote replication of the results and to enable further research.

We used the methods introduced by Straka et al. (2018) as a baselines, namely *First*, *Random*, *TextRank*, *clf-rf*, and *t2t*. We selected a task to create a headline out of the text of the article, which can be considered as a single sentence summary.

We proposed an extractive approach *Named Entity Density* that selects a sentence with the highest ratio between the number of entities and length of the sentence as the summary of the article. The experiments showed that *Named Entity Density* achieved nearly as good results as baseline selecting the first sentence of the article, which is a very hard baseline to beat, especially in the domain of news articles. Nevertheless, the summaries generated by *Named Entity Density* demonstrated that the selected sentences reflect the style of reports concisely identifying *to whom*, *when*, *where*, and *what* happened. We proposed using a combination of *Named Entity Density* and *First* summaries in voice applications. The voice application can initially present the first sentence of the article, and continue by follow-up created by *Named Entity Density* if a user requests additional information.

Next, we proposed two abstractive approaches based on the Seq2Seq architecture. The first approach, *Seq2Seq*, generates novel summaries using only tokens of the ar-

Method	Headline
Gold	Havlova Asanace by sama asanaci potřebovala *Havel's Asanace itself would need sanitation*
Seq2Seq	Havel se vrátil do divadla *Havel returned to the theater*
Seq2Seq–NER	Havel se s s Havlem Na zábradlí. Na hradě *Havel with with Havel at Na zábradlí. On Castle*
Gold	Hrad Bouzov nadchne cyklisty i zájemce o mučení a draky *Bouzov Castle will delight cyclists and those interested in torture and dragons*
Seq2Seq	Hrady a zámky na hrad. Kde se můžete vidět i na hrad *Castles and chateaux for the castle. Where you can yourself and castle*
SeqSseq–NER	Na kole na hrad *By bike to the castle*
Gold	Filmy z Indie opět v Praze *Films from India again in Prague*
Seq2Seq	V Indii se chystá na film o lásce *There are preparations for a movie about love in India*
Seq2Seq–NER	V Indii se vrací do Indie *In India, he returns to India*

Table 9. Examples of summarizations from the OOD test set

ticle's text. The second approach, *Seq2Seq–NER*, additionally uses the named entity labels of each word as its input. Experiments showed that both proposed methods achieve better results than the methods proposed previously by Straka et al. (2018). *Seq2Seq–NER* improved the results over *Seq2Seq* in automatic evaluation. This result demonstrated the usefulness of named entity labels for summarization. Furthermore, the results of the methods showed similar trends even on the out-of-domain test set.

Finally, we proposed a new metric, ROGUE$_{NE}$, that measures the overlap of named entities between the true and generated summaries. The results show that the current state-of-the-art methods struggle with named entities in summarization, and there is a significant opportunity for further research.

Acknowledgements

Bibliography

Allahyari, Mehdi, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D Trippe, Juan B Gutierrez, and Krys Kochut. Text summarization techniques: a brief survey. *arXiv preprint arXiv:1707.02268*, 2017. doi: 10.14569/IJACSA.2017.081052.

Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014. doi: 10. 3115/v1/D14-1179.

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

Filatova, Elena and Vasileios Hatzivassiloglou. Event-based extractive summarization. 2004.

Finch, Sarah E, James D Finch, Ali Ahmadvand, Xiangjue Dong, Ruixiang Qi, Harshita Sahijwani, Sergey Volokhin, Zihan Wang, Zihao Wang, Jinho D Choi, et al. Emora: An inquisitive social chatbot who cares for you. *arXiv preprint arXiv:2009.04617*, 2020.

Hassel, Martin. Exploitation of named entities in automatic text summarization for swedish. In *NODALIDA'03–14th Nordic Conference on Computational Linguistics, Reykjavik, Iceland, May 30–31 2003*, page 9, 2003.

Hatzivassiloglou, Vasileios and Elena Filatova. Domain-independent detection, extraction, and labeling of atomic events. 2003.

Hochreiter, Sepp and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spaCy: Industrial-strength Natural Language Processing in Python, 2020. URL `https://doi.org/10.5281/zenodo.1212303`.

Jabeen, Saima, Sajid Shah, and Asma Latif. Named entity recognition and normalization in tweets towards text summarization. In *Eighth International Conference on Digital Information Management* (*ICDIM 2013*), pages 223–227. IEEE, 2013. doi: 10.1109/ICDIM.2013.6694007.

Kågebäck, Mikael, Olof Mogren, Nina Tahmasebi, and Devdatt Dubhashi. Extractive summarization using continuous vector space models. In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality* (*CVSC*), pages 31–39, 2014.

Khademi, Mohammad Ebrahim and Mohammad Fakhredanesh. Persian automatic text summarization based on Named Entity Recognition. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, pages 1–12, 2020. doi: 10.1007/s40998-020-00352-2.

Konkol, Michal, Miloslav Konopík, Magda Ševčíková, Zdeněk Žabokrtský, Jana Straková, and Milan Straka. CoNLL-based Extended Czech Named Entity Corpus 2.0, 2014. URL `http://hdl.handle.net/11234/1-3493`. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Lin, Chin-Yew. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81, 2004.

Liu, Linqing, Yao Lu, Min Yang, Qiang Qu, Jia Zhu, and Hongyan Li. Generative adversarial network for abstractive text summarization. *arXiv preprint arXiv:1711.09357*, 2017.

Liu, Yang. Fine-tune BERT for extractive summarization. *arXiv preprint arXiv:1903.10318*, 2019.

Liu, Yang and Mirella Lapata. Text summarization with pretrained encoders. *arXiv preprint arXiv:1908.08345*, 2019. doi: 10.18653/v1/D19-1387.

Luong, Minh-Thang, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015. doi: 10.18653/v1/D15-1166.

Marek, Petr and Štěpán Müller. SumeCzech-NER, 2021. URL `http://hdl.handle.net/11234/1-3505`. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Mihalcea, Rada and Paul Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 404–411, 2004.

Müller, Štěpán. Named Entity Recognition. Research project, CTU in Prague, 2020a.

Müller, Štěpán. Text Summarization Using Named Entity Recognition. B.S. thesis, CTU in Prague, 2020b.

Nallapati, Ramesh, Bowen Zhou, Caglar Gulcehre, Bing Xiang, et al. Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*, 2016. doi: 10.18653/v1/K16-1028.

Nenkova, Ani. Automatic text summarization of newswire: Lessons learned from the document understanding conference. 2005.

Nobata, Chikashi, Satoshi Sekine, Hitoshi Isahara, and Ralph Grishman. Summarization System Integrated with Named Entity Tagging and IE pattern Discovery. In *LREC*, 2002.

Pal, Alok Ranjan and Diganta Saha. An approach to automatic text summarization using Word-Net. In *2014 IEEE International Advance Computing Conference (IACC)*, pages 1169–1173. IEEE, 2014. doi: 10.1109/IAdCC.2014.6779492.

Pichl, Jan, Petr Marek, Jakub Konrád, Petr Lorenc, Van Duy Ta, and Jan Šedivý. Alquist 3.0: Alexa Prize Bot Using Conversational Knowledge Graph. *arXiv preprint arXiv:2011.03261*, 2020.

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

Schulze, Frederik and Mariana Neves. Entity-supported summarization of biomedical abstracts. In *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016)*, pages 40–49, 2016.

Ševčíková, Magda, Zdeněk Žabokrtský, Jana Straková, and Milan Straka. Czech Named Entity Corpus 2.0, 2014. URL `http://hdl.handle.net/11858/00-097C-0000-0023-1B22-8`. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Song, Shengli, Haitao Huang, and Tongxiao Ruan. Abstractive text summarization using LSTM-CNN based deep learning. *Multimedia Tools and Applications*, 78(1):857–875, 2019. doi: 10.1007/s11042-018-5749-3.

Straka, Milan, Nikita Mediankin, Tom Kocmi, Zdeněk Žabokrtský, Vojtěch Hudeček, and Jan Hajič. Sumeczech: Large Czech news-based summarization dataset. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

Straková, Jana, Milan Straka, and Jan Hajič. Neural Architectures for Nested NER through Linearization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5326–5331, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1527. URL `https://www.aclweb.org/anthology/P19-1527`.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

Weidman, Seth. Language Translation with TorchText, 2019. URL `https://pytorch.org/tutorials/beginner/torchtext_translation_tutorial.html`.

Yao, Kaichun, Libo Zhang, Dawei Du, Tiejian Luo, Lili Tao, and Yanjun Wu. Dual encoding for abstractive text summarization. *IEEE transactions on cybernetics*, 2018. doi: 10.1109/TCYB.2018.2876317.

Zhang, Yong, Joo Er Meng, and Mahardhika Pratama. Extractive document summarization based on convolutional neural networks. In *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*, pages 918–922. IEEE, 2016. doi: 10.1109/IECON.2016.7793761.

Address for correspondence:
Petr Marek
`marekp17@fel.cvut.cz`
Faculty of Electrical Engineering
Czech Technical University in Prague
Technická 2
166 27 Praha 6 - Dejvice, Czech Republic