Institute for Natural Language Processing

University of Stuttgart
Pfaffenwaldring 5 b
D–70569 Stuttgart

Bachelorarbeit

Automatic recognition of structures in obituaries

Valentino Sabbatino

Course of Study:	Softwaretechnik
Examiner:	Dr. Roman Klinger, Prof. Dr. Sebastian Padó
Supervisor:	Laura-Ana-Maria Bostan, M. Sc.
Commenced:	November 15, 2018
Completed:	May 15, 2019

Abstract

Obituaries are a less common text type in research that contains a lot of information about people, events in history and culture. The information that can be obtained by zoning such obituaries enables new research, e.g., in social studies. Our work focuses on the question if the structuring of obituaries is possible and viable. Therefore we created a corpus for this work containing 20058 obituaries of which 1008 were annotated manually by us. We implemented four models, a CNN text classifier and three variations of a Bi-LSTM sequence labeler, to see if the zoning procedure is possible and which among the models performs best for this task. The CNN text classifier showed the most promising results together with the variant of the Bi-LSTM model using a Bag-of-Word model.

Contents

List of Figures

List of Tables

List of Abbreviations

BOW Bag-Of-Word. 15

CNN Convolutional Neural Network. 15

CRF Conditional Random Field. 17

LSTM Long Short Term Memory. 9

NER Named Entity Recognition. 16

NLP Natural Language Processing. 16

PoS Part of Speech. 16

ReLU Rectified Linear Unite. 30

RNN Recurrent Neural Network. 17

SLP Sequence Labeling Problems. 16

TC Text Classification. 15

1 Introduction

Text classification is an important task in natural language processing and describes the assignment of documents to one or more predefined classes [13]. Some popular areas include classifying news as politics, sports, lifestyle, etc., movie reviews as good, neutral, or bad, or jokes as funny or not funny [19]. Text Classification can also be applied in many other areas, such as "web search, information retrieval, ranking and document classification" [15], spam filtering, author identification, and sentiment analysis [21]. Previous work dealt with a variety of text genres, like scientific articles, news reports, movie reviews, and advertisements [12].

This work focuses on obituaries as a lesser used text type in text classification research. An obituary is a notice of a death, typically found in newspapers. It informs about the recent death of a person. It usually includes a brief biography of the deceased person, where styles, formats, and information presented therein vary slightly from culture to culture [2]. Further, it contains information on the living family members and information about the upcoming funeral, such as visitation, burial service, and memorial information. The cause of death is also usually given in an obituary [2]. So each obituary contains certain recurring elements that we aim to identify to answer the question of whether one can predict the structures in an obituary. The approach to categorize each sentence into such categories can be formalized as a sequence labeling problem, where we focus on the following sections in an obituary: Personal information (including names of the deceased, birth date,date of death, and cause of death), biographical sketch, tribute, family, and funeral information, such as time, place, and date of the funeral [2]. Sequence labeling is a type of pattern recognition task, similar to text classification, and consists of assigning a sequence of class labels to a sequence of inputs, e.g., the sentences in an obituary [20]. Other fields where sequence labeling is applied are genomic research, health informatics, and anomaly/ intrusion detection [23]. Through automatic structure recognition in obituaries, we enable further research in this field, e.g. the investigation of a connection between work and cause of death, the investigation of the portrayal of war heroes, and the analysis of linguistic, structural or cultural differences in obituaries.

To realize our goal we implement a Convolutional Neural Network text classifier as a baseline, which does not use sequences as features. With this, we have a base model to examine how well we can structure obituaries automatically. To see if the context can improve the structuring process we implement further models, which realize a sequence labeler. With this, we include positional information of the sentences into the structuring process. Through a comparison of our models, we will not only determine which model performs best but can also determine where each model has its advantages and weaknesses. The main contribution of this work is the annotated collection of obituaries, a Convolutional Neural Network classifier and the neural sequence labeler for structure recognition. Furthermore, we will illustrate the influence and difference of different sequence labeling approaches and show which model performs best.

In the following section, the foundations (2) and related works (3) are presented. The foundations cover the topics Text Classification (2.1), Sequence Labeling (2.2), and Zoning of documents (2.3). In the section related work (3) we cover works that dealt with obituaries in a computer science context. In the subsequent Sectionr 4, we present our dataset, explain the data selection procedure (4.1), the annotation procedure (4.2), and analyze the dataset in Section 4.3. The Section 5 consists of the explanation of the experimental setup (5.1), the presentation of the results (5.2), and a discussion (5.3). At the end comes the conclusion in Section 6.

2 Foundations

This section covers the topics of text classification (Section 2.1), sequence labeling (Section 2.2), and the zoning of documents (Section 2.3), giving an introduction into each topic and giving an overview of the methods we will use in later sections.

2.1 Text Classification

With the increase of documents in electronic form, the need arises to organize them, which can be made possible by Text Classification (TC), which is a type of pattern recognition task [21]. Uysal [21] stated that TC is applied to a variety of domains, such as topic detection, spam filtering, author identification, web page classification, and sentiment analysis. The goal of TC is to assign one (or more) predefined classes to a document [13]. To obtain that goal supervised learning techniques that learn a classifier through pre-labeled examples, the so-called training data set, are needed [13, 19]. As a result a classification function (f) is learned, which maps a document (d) to a class (C),i.e:

$$f : d \to C.$$

The size of the training data set and the testing data set (unlabeled documents) is a decisive factor for the accuracy of the classifier. The provision of a small number of training data could lead to an inaccurate classifier because it lacks substantial knowledge. On the other hand, if the number of training data is too large, it leads to a problem called "Overfitting" [19]. Sriram et al. [19] described "Overfitting" as a degradation of a classifiers' performance, caused by an over adaption due to a large number of used training data. If a learning model excessively pursues maximizing training accuracy, it can learn a very complex model. However, there is a risk that it will fall into overfitting. The problem here is that an overfitting model may memorize non-predictive features of the training data, instead of learning to generalize from a trend [6].

The documents with which a classifier has to work, need to be transformed into a representation suitable for the learning algorithm. Documents are typically strings of characters [13]. To represent the document in a structured manner the Bag-Of-Word (BOW) model is commonly applied. The technique splits the text into words, where each word represents a feature. Worth mentioning is, that the model ignores the exact order in which a word occurs [19]. Examples of classification techniques are Naïve Bayes, Support Vector Machine, and Convolutional Neural Network (CNN) classifier.

2.1.1 Convolutional Neural Network

CNNs are neuronal networks that make use of the internal structure of the data, e.g., "the 2D structure of image data through convolution layers, where each computation unit responds to a small region of input data (e.g., a small square of a large image)"[14]. That enables us to make use of the word order in documents for text classification. Thus, CNNs found application in entity search, sentence modeling, word embedding learning, and product feature mining. The first layer of a CNN is the input layer, in which words in sentences are converted into word vectors by table lookup. Thereby, the word vectors are “either trained as part of CNN training, or fixed to those learned by some other method [...] from an additional large corpus”[14]. Word2Vec is one such method to learn word embeddings. In our work, we make use of a pre-trained Word2Vec model, which is further explained in Section 5. On the input layer follow the convolutional layers, which “consists of several computation units, each of which takes as input a region vector that represents a small region of the input image and applies a non-linear function to it” [14]. The output of the convolutional layers goes into pooling layers, where the dimensionality of the previous output is reduced and a small degree of translational invariance is conferred into the model [22]. Commonly-used methods for pooling are max-pooling and average-pooling. The last layer is the output layer. Figure 2.1 illustrates how a CNN is built up, showing the input, convolutional layer, pooling layer, and output layer.

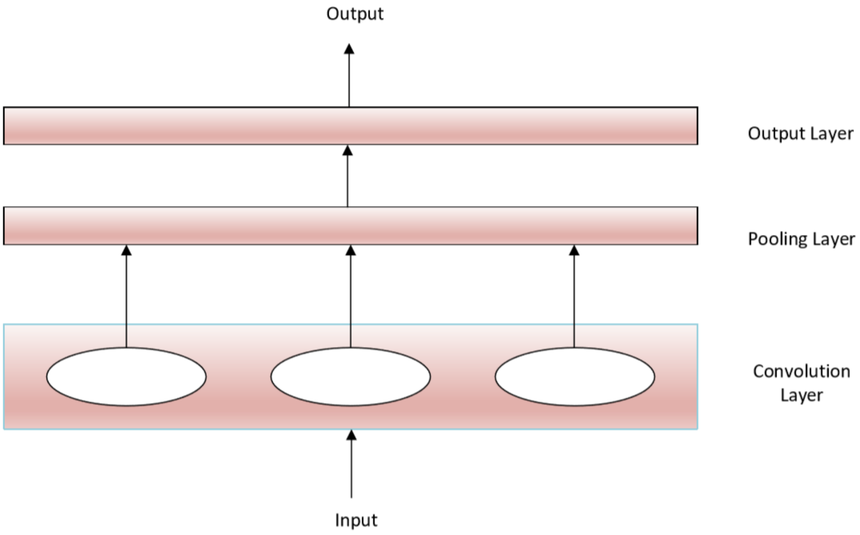


Figure 2.1: Convolutional neural network

2.2 Sequence Labeling

Part of Speech (PoS) tagging, chunking, or Named Entity Recognition (NER), are some examples of Natural Language Processing (NLP) tasks, which are commonly modeled as Sequence Labeling Problems (SLP) [20]. Sequence labeling has a very wide range of applications, such as classifying protein sequences into existing categories to learn the functions of a new protein, classifying query log sequences to distinguish web-robots from human users, and classifying transaction sequence data in a bank for the purpose of combating money laundering [23]. Sequence labeling is a type of pattern recognition task and consists of assigning a sequence of class labels $\vec{y} = (y_1, ..., y_n) \in Y^n$ to a sequence of inputs $\vec{x} = (x_1, ..., x_n) \in X^n$ [20]. Words in a sentence correspond to tokens x_i in the input sequence $\vec{x}$, which are mapped to class labels $\vec{y}$ [17, 20], which e.g., can be POS tags or entity classes [20]. Settles and Craven [17] used Conditional Random Fields, statistical graphical models, that showed "state-of-the-art accuracy on virtually all of the sequence labeling tasks".

2.2.1 Conditional Random Fields

Conditional Random Field (CRF) are a probabilistic framework for labeling structured data and define conditional probability distributions $P_{\vec{\lambda}}(\vec{y}|\vec{x})$ of label sequences given input sequences [18, 20]. The CRFs we look at are a special form for sequential data called linear-chain CRFs [20]. The most probable label sequence for input sequence $\vec{x}$ is

$$\hat{y} = \underset{\vec{y} \in Y^n}{\arg\max}\ P_{\vec{\lambda}}(\vec{y}|\vec{x})$$

and can be computed using the Viterbi algorithm [20]. The conditional probability distributions $P_{\vec{\lambda}}(\vec{y}|\vec{x})$ can be calculated as follows:

$$P_{\vec{\lambda}}(\vec{y}|\vec{x}) = \frac{1}{Z_{\vec{\lambda}}(\vec{x})} \cdot \exp(\sum_{j=1}^{n} \sum_{i=1}^{m} \lambda_i f_i(y_{j-1}, y_j, \vec{x}, j)),$$

where $f_i(y_{j-1}, y_j, \vec{x}, j)$ is a feature function, with the inputs $\vec{x}$ (e.g., a sentence), the position j in the sequence $\vec{x}$, the label of the current word y_j, and the label of the previous word y_{j-1} [4], feature weights λ_i, and $Z_{\vec{\lambda}}(\vec{x})$ as a normalization factor over all possible labelings of $\vec{x}$ [17]. $Z_{\vec{\lambda}}(\vec{x})$ can be calculated as follows:

$$Z_{\vec{\lambda}}(\vec{x}) = \sum_{y'} \exp(\sum_{j=1}^{n} \sum_{i=1}^{m} \lambda_i f_i(y'_{j-1}, y'_j, \vec{x}, j)).$$

2.2.2 Bi-LSTM

"A Recurrent Neural Network (RNN) maintains a memory based on history information, which enables the model to predict the current output conditioned on long distance features" [11]. A RNN follows the structure seen in Figure BLA, where we have first of all an input layer, followed by a hidden layer, and an output layer corresponding to the predicted labels. The input layer represents features in form of one-hot-encoding or word embeddings at a time t. LSTM is a RNN architecture that is better at finding and exploiting long-range dependencies. During sequence classification, we have access to the past and future input features for a given time t. Making use of that results in a bidirectional LSTM network [11] illustrated in Figure 2.2.

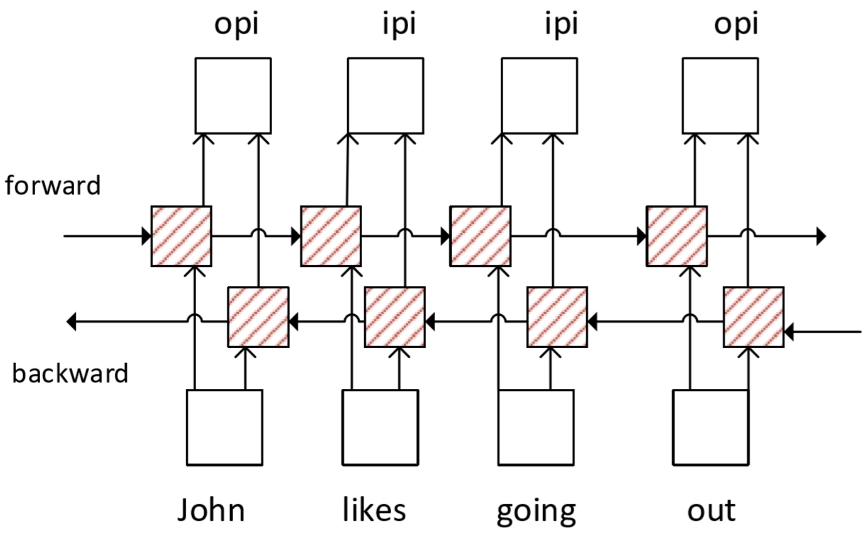


Figure 2.2: Three layers of a Bi-LSTM model. Input, hidden, and output layer.

2.3 Zoning of documents

Many NLP tasks focus on the extraction of specific types of information in documents. To make searching and retrieving information in documents accessible, the logical structure of documents in titles, headings, sections, and thematically related parts must be recognized. Given the number of documents available in electronic form, this task needs to be done automatically [16]. For example, a notable amount of work focuses on the zoning of scientific documents. Guo et al. [8] stated that readers of scientific work may be looking for "information about the objective of the study in question, the methods used in the study, the results obtained, or the conclusions drawn by authors". Various methods have been proposed, that deal with the zoning of scientific literature. Some methods use section names typically seen in scientific documents, make use of argumentative zoning, qualitative dimensions, or conceptual structure of documents [8]. The recognition of document structures makes generally use of two sources of information. One is the text layout. This makes it possible to recognize relationships between the various structural units such as headings, body text, references, figures, etc. In addition to the structure of the text, the wording and content itself can be used to recognize the connections and semantics of text passages. So, the structuring of documents based on the sequence of text objects and layout features can be formulated as a classification problem which may be solved with a machine learning method [16].

3 Related Work

Obituaries have been the subject of various works in a variety of fields. For example, Herat [10] investigated how certain language expressions are used in obituaries in Sri Lanka, how religion and culture play a role, and how language reflects social status. In Epstein and Epstein [5] work they studied the relationship between career success, terminal disease frequency, and longevity using New York Times obituaries. These works by Epstein and Epstein [5] and Herat [10] represent the research in the medical and linguistic field. In this work we, want to focus on previous work that worked with obituaries in a computer science context. Most papers working with obituaries, which are presented here, focused on information extraction from various sources including obituaries.

Ford et al. [7] mined text data from obituary websites, with the intention to use it to prevent identity theft, where someone assumes the identity of a recently deceased person. Ford et al. [7] conducted a study, where the goal was to evaluate how "often and how accurately name and address fragments extracted from these notices developed into complete name and address information corresponding to the deceased individual". In their case study, they have discovered five obituaries websites and divided them into three categories. The first, category A, contained formatted pages on which the obituary messages were collected programmatically. The data was then extracted using a low-loss pattern-matcher. Category B websites were usually unformatted. Therefore, a special text analysis technique was required to extract the relevant data. Category C included websites including newsgroups, forums, and mailing lists, etc. The system they built focused on category A only. The extracted data is usually not complete enough for the anti-fraud application because it lacks a complete address. To solve the problem, they used a trusted information knowledge base, called Index, with name and address information. They then extracted the name and address fragments from the text and matched them against the Trusted Index to create a set of name and address candidates where each candidate matches one or more of the data fragments. Because the found set contained uncertainties, each record in the set was assigned a confidence level between 0 and 100 inclusive. The result set was then compared to "the Authoritative Source in order to determine which of the candidate records actually corresponded to the name and address of an individual reported as deceased, and to correlate these findings with the assigned confidence level" Ford et al. [7]. The results showed that the candidate records with confidence level 95 were found in about 80% of the cases in the Authoritative Source. This means that the records with the names, addresses, and cities matched the stored data index and only 80% of the records matched the Authoritative Source. The other 20% of the records did not appear in Authoritative Source. Their conclusion was that additional attributes, if available, can help refine the values assigned as confidence [7].

Alfano et al. [1] mined obituaries, which were collected from various newspapers, to get a better understanding of people's values. Their goal was to provide those values in a holistic picture of the virtues expressed by each community, presented in the form of network graphics that simultaneously represent hundreds or thousands of relationships in a single, easily digestible image. Alfano et al. [1] conducted three studies in which they used hand coding in the first and second studies, in which the obituaries were carefully read and labeled. In the third study, they further developed the results of previous studies with a semiautomatic, large-scale semantic analysis of several thousand obituaries. The first study made use of local obituaries, the second one of obituaries of famous people from the New York Times, and in the third study, they performed large-scale data-mining of local obituaries [1].

In his work "Personal Information Extraction from Korean Obituaries", Han [9] conducted a study which found an effective method for extracting various facts about persons from obituary web pages. Because of the similarities in description styles, Han [9] used a feature scoring method that applies prior knowledge in a heuristic way. His method achieved high performance for each attribute (e.g., Person Name, Affiliation, Position (Occupation), Age, Gender, Death Cause, etc.) based on recall and precision. The extraction results can be used for “extending existing biographical dictionaries and for acting as a seed dictionary for a named entity recognition”[9].

Xu and Embley [24] proposed an approach based on information extraction ontologies that extracts the expected ontological vocabulary and instance data from given web pages. After extracting the data from the web pages, they used machine-learned rules to determine if a web page contains interesting elements. Xu and Embley [24] have experimented with four applications: Car ads, obituaries, real estate houses and faculty pages. Their Goal was the Evaluation of the categorization performance over several types of web pages for real applications. Their results showed that recognition F-measures for all four applications were above 90% [24].

Bamman and Smith [3] have presented in their work a general, unsupervised model for learning life event classes from biographical texts along with the structure that connects them. Furthermore, by using this method to learn event classes from Wikipedia, they discovered evidence of systematic bias in the presentation of male and female biographies in which female biographies placed a significantly disproportionate emphasis on the personal events of marriage and divorce. They extracted 242,970 biographies from Wikipedia and applied their method to them. This work is of interest because it deals with deep biographical information (Wikipedia biographies), of which obituaries are also a part. In a quantitative analysis, the model they presented exceeded a strong baseline regarding the task of event time prediction, which also showed a qualitative distinction in the content of the biographies of men and women on Wikipedia.

4 Dataset

This section will explain the corpus we used to answer the question if we can predict the structure of an obituary. We will explain how our data was selected (Section 4.1), what our criteria for the annotation procedure were (Section 4.2), which include a statistical analysis (Section 4.2.1) of our corpus. In the Section 4.3 we give a further analysis of our annotated corpus.

4.1 Data Selection

Obituaries inform about the death of a person and provide the reader in general with a short biography of the deceased person, information about the surviving family members, and information about the upcoming funeral, such as visitation, burial service, and memorial information. The first sentences of an obituary include in most cases the place of death and the cause of death. The structure and contents can vary depending on the cultural background [2]. Obituaries can be differentiated into death notices and memorial advertisements, where a death notice omits most of the biographical information and a memorial advertisement is usually written by the family or a paid death ad writer. Obituaries are usually published in newspaper and online on designated websites or on websites of newspapers.

For the task of structure recognition in obituaries, we collected the obituaries from three websites that allowed our crawler access to them. The three websites are The Daily Item[1], where obituaries from the U.S.A. are published, Remembering.CA[2], which covers obituaries from Canada, and The London Free Press[3], which covers obituaries from London. From The Daily Item we collected 9975 obituaries, from Rembering.CA we collected 9984 obituaries, and from The London Free Press, we collected 99 obituaries. So in total, we have 20058 obituaries in our dataset. Table 4.1 provides some examples of sentences from obituaries with the corresponding given labels.

Class	Sentence
personal information	Helen Jarrett, of North Fort Myers, Fla., passed away on Nov. 26, 2018, at the Page Rehabilation Hospital, Fort Myers.
biographical sketch	She was born May 16, 1940, in Bloomsburg, a daughter of the late Sheldon and Althea (Hartzel) Bucher of Bloomsburg.
characteristics	Mildred enjoyed cooking and puzzles.

Table 4.1: Examples of sentences from an obituary

[1] The Daily Item: http://obituaries.dailyitem.com/obituaries/all-categories/search/

[2] Remembering.CA: http://www.remembering.ca/obituaries/all-categories/search/

[3] The London Free Press: http://lfpress.remembering.ca/

4.2 Annotation Procedure

In each obituary, we can find certain recurring elements, such as personal information (including names of the deceased, birth date, date of death, and cause of death), biographical sketch, tribute, family, and funeral information, such as time, place, and date of the funeral [2]. Resulting from these recurring elements we use eight labels in total, which consists of personal information (pi), biographical sketch (bs), characteristics (c), tribute (t), expression of gratitude (g), family (f), funeral information (fi), and other (o) to structure an obituary. From those 20058 obituaries, we chose 1008 randomly and labeled them using the annotation guideline. The labels will be explained in the following sections, to clarify when certain labels were given. We also provide examples of sentences from obituaries to support these explanations where we have changed the names of the affected ones.

Personal Information

The personal information label is a subset of the more general biographical sketch label and serves more the purpose to classify most of the introductory clauses in obituaries. We have chosen to refer to a sentence as personal information when one or more of these points apply:

- Includes the name of the deceased
- Includes the date of death
- Includes the cause of death
- Includes the place of death

Example: “John Doe, 64, of Newport, found eternal rest on Nov. 22, 2018.”

Biographical sketch

The biographical sketch is similar to a curriculum vitae. Sections in a person’s life fall into this category. However, it should not be regarded exclusively as a curriculum vitae, since it forms the superset of personal information. We decided to label a sentence as biographical sketch if one or more of these points apply:

- Includes the place of birth
- Includes the birth date
- Includes the last place of residence
- Includes the wedding date
- Includes the duration of the marriage
- Includes the attended schools
- Includes the occupations
- Includes the further events in life

Example: “He entered Bloomsburg State Teachers College in 1955 and graduated in 1959.”

Characteristics

The characteristics are a class, which are recognizable by the fact that the person himself is described. They can be character traits or things the deceased loved to do. Apart from hobbies, the deceased faith is also part of the characteristics.

Example: “He enjoyed playing basketball, tennis, golf and Lyon’s softball.”

Tribute

Sentences are labeled as tribute if it is about a major achievement of the deceased. We consider mainly mentions of contributions to society as a tribute.

Example: “His work was a credit to the Ukrainian community, elevating the efforts of its arts sector beyond its own expectations.”

Expression of gratitude

Sentences in obituaries are labeled as an expression of gratitude if any form of gratitude occurs in it, be it directed to doctors, friends, or other people. In most cases, it comes from the deceased’s family.

Example: “We like to thank Leamington Hospital ICU staff, Windsor Regional Hospital ICU staff and Trillium for all your great care and support.”

Family

The label family is given to all sentences that address the survivors or in which previously deceased close relatives, such as siblings or partners, are mentioned. The mentioning of the wedding date is not covered by this category, because we considered it an event and as such, it fell under the biographical sketch category. But if the precedence of those persons is mentioned it falls in this category. If a marriage is mentioned without the wedding date or the duration it falls into the family category.

Example: “Magnus is survived by his daughter Marlene (Dwight), son Kelvin (Patricia), brother Otto (Jean) and also by numerous grandchildren & great grandchildren, nieces and nephews.”

Funeral information

Sentences are labeled as funeral information when it deals with information related to the funeral, such as date of the funeral, time of the funeral, place of the funeral, and where to make memorial contributions.

Example: “A Celebration of Life will be held at the Maple Ridge Legion 12101-224th Street, Maple Ridge Saturday December 8, 2018 from 1 to 3 p.m.”

Other

Everything that does not fall into the above-mentioned classes is classified as other.

Example: “Dad referred to Lynda as his Swiss Army wife.”

4.2.1 Statistical Analysis

As above mentioned the dataset consists of 20058 obituaries from three different sources. The labeled dataset, consisting of 1008 obituaries, consists of 11087 sentences. The maximum length of a sentence, measured in the number of words, was 321. From the 1008 obituaries are 475 from the website *The Daily Item*, therefore they are labeled with USA in Table 4.2 and Figure 4.2. 445 labeled obituaries are from the website *Remembering.CA* and labeled with Canada in Table 4.2 and 4.2. 88 labeled obituaries are from *The London Free Press* and are labeled with UK in Table 4.2 and Figure 4.2.

Most sentences in the dataset are labeled as *biographical sketch*, followed by *funeral information* with 2831 sentences. The third most assigned label is *family* with a number of 2195. The least assigned label is *tribute*, with 11 sentences, followed by *gratitude* with 144 sentences. The distribution of the number of given labels is visualized in Figure 4.1 and can also be looked up in the column *total* in Table 4.2. The number of sentences per subset, USA, Canada, and UK, can be seen in the corresponding columns.

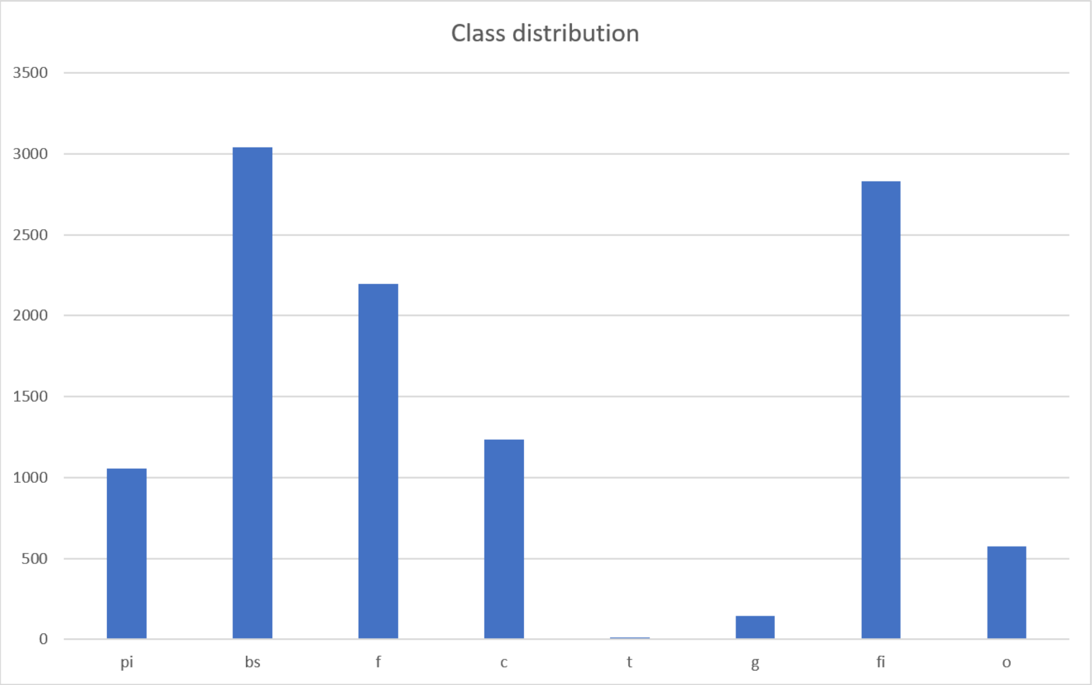
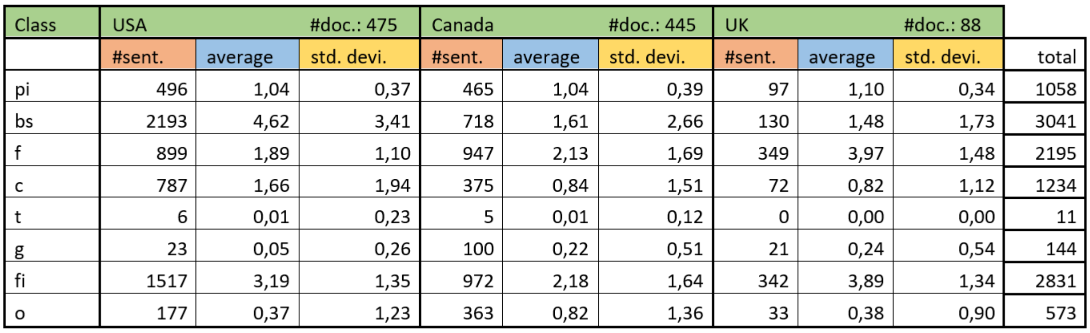


Figure 4.1: Total label distribution in the dataset
pi: personla information, bs: biographical sketch, f: family, c: characteristics, t: tribute, g: gratitude, fi: funeral information, o: other

What can be observed in Figure4.2 is that the label *pi* was assigned on average 1.04 for the obituaries from USA and Canada. The obituaries from the UK had the label *pi* assigned 1.10 times on average, which shows that those numbers are very close together. The standard deviation shows also that the deviation is small. For the class *biographical sketch*, there is some discrepancy compared to the other classes. The mean for the obituaries from the USA is a lot higher, with a value of $4,62$ than for Canada and the UK, with values of 1.61 and 1.48. This can be interpreted in such a way that the writers of the obituaries in the USA put more value on the biography of a person than in the other countries, but the standard deviation suggests, that the number of sentences labeled as biographical sketch is spread out over a wider range. The mean for Canada and the UK are close, but the standard deviation of the obituaries from Canada, with a value of 2.66, is higher than for the ones from the UK. For the class *family* we can see that the label was assigned on average 3.97 for the obituaries from the UK, which is higher than the values from USA and Canada, which have the corresponding values of 1.89 and 2.13. This means that they have dedicated more sentences about the family members in the obituaries from the UK than in the other two. However, one must bear in mind that the number of obituaries from the UK is much lower than that from the USA and Canada. The average for the number of sentences labeled as *characteristics* per obituary from the USA is almost twice as big as the values for Canada and UK, however, the standard deviation is also higher than for the other two. The label *tribute* is the least given class so in the whole dataset it occurs only eleven times. So the average is only 0.01 for obituaries from USA and Canada and 0 for the ones from the UK because no sentence from that subset was labeled as a tribute. The label *gratitude* is the second most uncommon class, where the average for the USA is lower, with a value of $0,05$ than for the other two, with values of $0,22$ and $0,24$. For the class *funeral information* we can deduce, that the obituaries from the UK give the most information about the funeral itself, followed by USA. The least information is given in obituaries from Canada with only an average value of 2.18. The last class is other. The label *other* was assigned on average $0,37$ and $0,38$ for the obituaries from the USA and UK, which are very close. Canada has the most sentences labeled as *other*, with a value of 0.82, in comparison with USA and Canada.



Class	USA		#doc.: 475	Canada		#doc.: 445	UK		#doc.: 88	
	#sent.	average	std. devi.	#sent.	average	std. devi.	#sent.	average	std. devi.	total
pi	496	1,04	0,37	465	1,04	0,39	97	1,10	0,34	1058
bs	2193	4,62	3,41	718	1,61	2,66	130	1,48	1,73	3041
f	899	1,89	1,10	947	2,13	1,69	349	3,97	1,48	2195
c	787	1,66	1,94	375	0,84	1,51	72	0,82	1,12	1234
t	6	0,01	0,23	5	0,01	0,12	0	0,00	0,00	11
g	23	0,05	0,26	100	0,22	0,51	21	0,24	0,54	144
fi	1517	3,19	1,35	972	2,18	1,64	342	3,89	1,34	2831
o	177	0,37	1,23	363	0,82	1,36	33	0,38	0,90	573

Table 4.2: Table with general information such as number of sentences, average distribution, and the standard deviation
pi: personla information, bs: biographical sketch, f: family, c: characteristics, t: tribute, g: gratitude, fi: funeral information, o: other

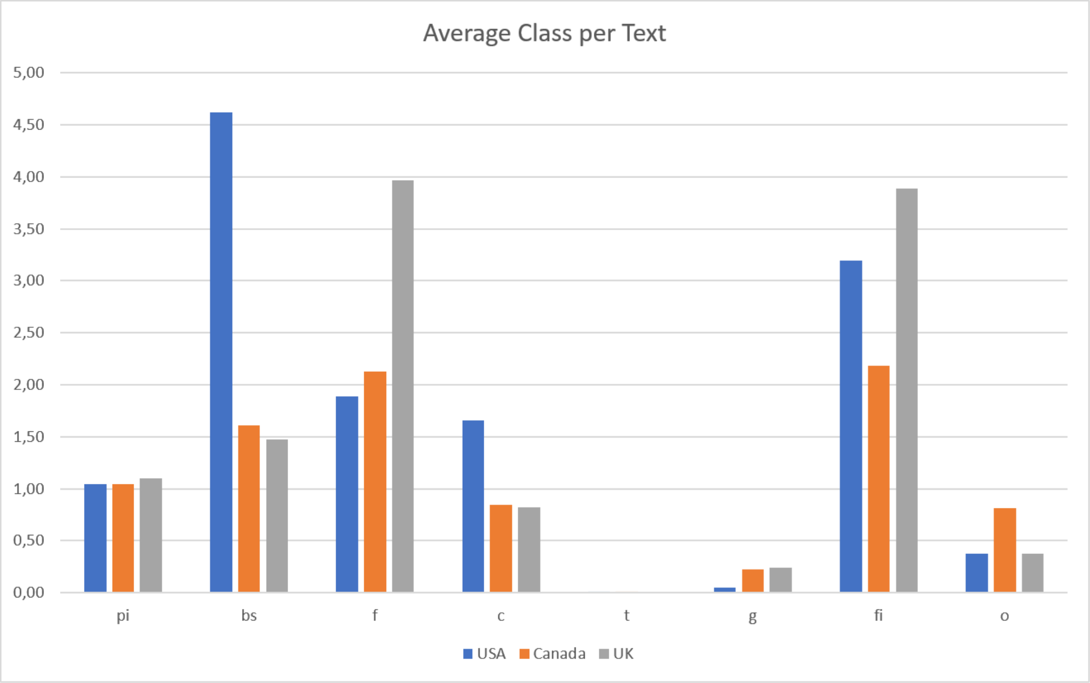


Figure 4.2: Average class assignment per obituary
pi: personla information, bs: biographical sketch, f: family, c: characteristics, t: tribute, g: gratitude, fi: funeral information, o: other

4.2.2 Annotators agreement

To evaluate the annotation guideline and to have a metric for it we calculated the annotator's agreement via Cohen's Kappa Score and Fleiss Kappa Score. We selected three annotators, which had to annotate the dataset consisting of the 1008 obituaries which are used in our experiment presented in chapter 5. The results of the inter annotators agreement can be seen in Table 4.3. The results for Cohen's Kappa between the first and the second annotator is 0.8598, which is a good value corresponding to "good agreement". For the first and third annotator, we have a value of 0.8678. Between the second and third annotator, we have a value of 0.8720. Because the Cohen's Kappa Score was greater than 0.81 we can interpret these results as "good agreement". The result of Fleiss Kappa supports this with a value of 0.8665.

After the annotation process, the results of the inter annotators agreement were discussed. Possible reasons for a deviation in some classifications may lay in the difficulties to classify some of the rarer cases that appeared, for example, the class *tribute*. Furthermore, there were some overlaps. For example the differentiation between the class *family* and *biographical sketch* was not easy because of the occurrence of a wedding date, which we considered an event, in connection with the other *family* criteria. Further existed overlaps regarding personal information, which was expected considering that *personal information* is a special case of *biographical sketch*.

Annotators	**A1 & A2**	**A1 & A3**	**A2 & A3**
Cohen Kappa Score	0.8598	0.8678	0.8720
Fleiss Kappa Score	0.8665		

Table 4.3: Annotators Agreement calculated with Cohen's Kappa and Fleiss's Kappa

4.3 Data Analysis

In subsection 4.2.1 we could see that there were already differences in the distribution of the classes between the countries. If we take a look at the class *biographical sketch* we can observe that that class is more present in obituaries from the USA than in the other. Through the annotation process, we could observe that in the obituaries from the USA they often emphasize the exercise of military service, which was less present in the obituaries from the other two countries, thus having more sentences dedicated to the biography. Another deviation regarding the others can be seen for the obituaries from the UK regarding the class *family*. We noticed during the annotation that the free time activities were often connected with family members, which could explain the increased assignment of the class *family*. Moreover, there is the problem that we had not many samples from the UK. The lack of obituaries from the UK could lead to a subselection which put more value on the families of the deceased. For the class *funeral information* we could see that obituaries from the UK had apparently fewer sentences about funeral information than the other two. We could not determine the cause of this. Aside from the information, we get from the statistical data in combination with a contextual examination of the obituaries we could observer further particularities of the dataset. By reading through the obituaries we noticed that depending on the sex the focus was different. Especially when the deceased was a female we could observe that the biographical sketch was more focused on the family achievements compared to males. Additionally, there was more focus on the role of a mother or grandmother.

5 Experiment

With this work, we want to examine if we can predict the structure of an obituary. As a baseline to predict the structures in obituaries we use a CNN text classifier. The CNN text classifier helps us in the recognition of the structure regarding each sentence. In the next step, we want to further investigate if positional information of the sentences in the obituary help in the structuring process. For that, we use three sequence labeler based on a bi-LSTM sequence classifier.

This chapter serves to explain the experiment structure (Section 5.1), which were used on the dataset described in Section 4.In addition to the settings, the results are also compared (in Section 5.2) with each other and discussed (Section 5.3).

5.1 Experimental setup

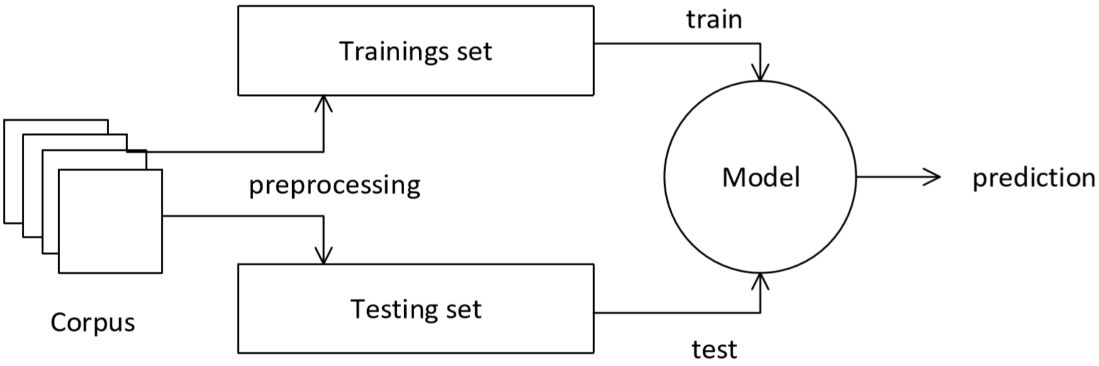


Figure 5.1: Abstract model of our experimental setup

Each experiment follows the same setup, which is depicted in Figure 5.1. At the beginning, we have our labeled corpus that goes through a preprocessing. After the preprocessing, we obtain a training and testing set. Our training set is used to train our models, which are then further used to predict the labels on our testing set.

This section serves to explain the preprocessing process (Section 5.1.1) and explains the implemented methods (Section 5.1.2).

5.1.1 Preprocessing

For the experiment, we used the labeled dataset described in chapter 4. For all classifiers we split the dataset into a training set, consisting of 70% of the previously shuffled obituaries, and a training set, consisting of the remaining 30%. The obituaries are then cleaned from punctuations. In the next step, we go through we training set and create a mapping where we map each word and label to an index. This is needed because the classifiers need the sentences in a format where each word is represented through an integer. Finally, to bring all sentences to the same length we use padding, where zeros are inserted to equalize.

5.1.2 Methods

For the first experiment step we used a CNN text classifier as a baseline and a bi-directional LSTM sequence labeler for comparison.

CNN Text Classifier

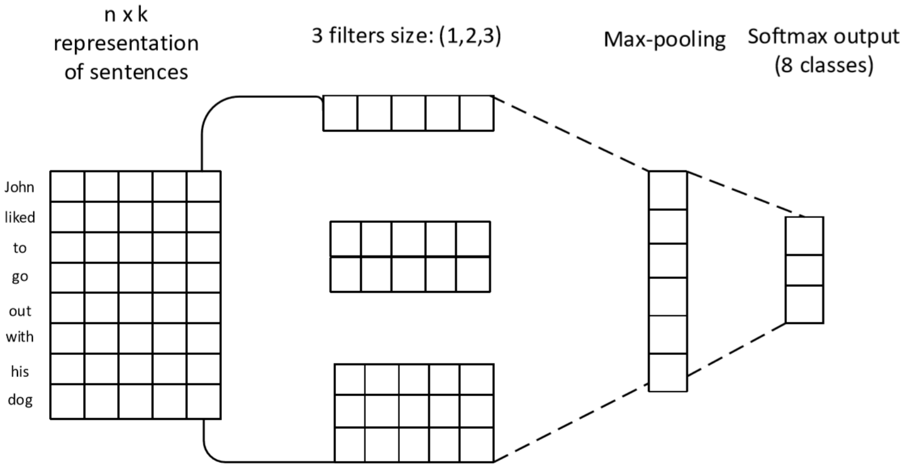


Figure 5.2: General setup of a CNN model

We used Keras[1] to construct our model. We used the “Sequential” model provided by Keras, which allows us to easily stack sequential and recurrent layers of the network in order from input to output. The input consists of the sentences in the training set. Our first Layer is a 2D convolutional layer with 128 output channels. We have a Kernel size of 1, which means we have a 1 × 1 filter matrix. As activation function we use Rectified

[1] Keras - https://keras.io/

Linear Unite (ReLU). Following is a 2D max pooling layer with a kernel of dimension ((max. length of a sentence) − (Kernel size) +1) × 1. We set a dimension which is always depended from the maximum sentences length and the Kernel size. Next, we add another 2D convolutional and max-pooling layer, where we have a 2 × 2 filter matrix. The third layer is another 2D convolutional and max-pooling layer with a Kernel size of 3. The output of the convolutional layers is then flattened. For the output layer, we use the dense layer, provided by Keras, which is the size of the number of our classes. As activation function we have soft-max. The CNN is outlined in Figure 5.2.

Bi-directional LSTM Sequence Labeler

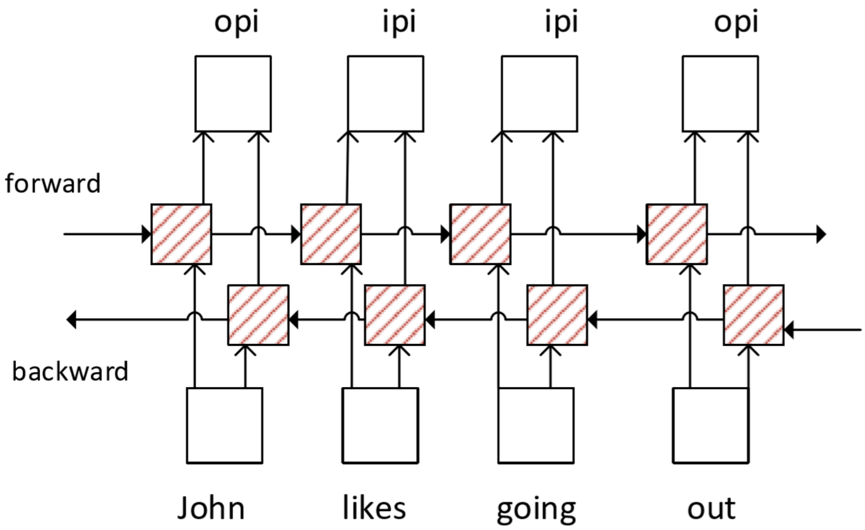


Figure 5.3: General setup of bidirectional LSTM model

For the bi-LSTM sequence labeler we have three variants consisting of a model using the BOW model, a model using pre-trained word embeddings, and a model that uses on top of the model with pre-trained word embeddings a CRF layer. The base structure of the Bi-LSTMs is shown in Figure 5.3, which is a repetition from Section 2.2.2. The sentences of an obituary are concatenated resulting in one sequence containing all sentences of the corresponding obituary. The labels corresponding to each sentence are changed into labels that correspond to each word. The new labels follow the BIO scheme resulting in 8 * 3 labels. The first layer is the embedding layer. For the word embeddings, we use the pre-trained Word2Vec Google News model[2], which is the only model we used for the pre-trained embeddings in all experiments. The Word2Vec Google News model consists of 3 million 300-dimensional word vectors. The next layer is the bi-LSTM layer. Therefore

[2] Google Word2Vec - https://code.google.com/archive/p/word2vec/

we have a bidirectional layer that functions as a wrapper in Keras which takes a LSTM layer as an argument, where the first hidden layer 100 memory units have. The output layer is a fully connected layer that outputs one value per timestep. For the output layer, we use a dense layer, which has a time distributed wrapper layer around. As the activation function, we use soft-max. The bi-LSTM using the CRF on top follows the same structure, with only changing the activation function of the dense layer to ReLU and having as output layer a CRF layer.

5.2 Results

To compare the results of each classifier we decided to use for each model the same parameters. This means, that each classifier operates with similar configurations to improve the comparability, but are not optimized. The number of epochs is set to 10. This number was chosen because most of the classifiers achieved acceptable results while also ensuring a shorter runtime. Our labeled corpus is split into a training set and test set, where 70% of the corpus is used for training and 30% for testing. The validation ration was set to 0.1. We used for this a built-in function of Keras, which takes the last fraction apart for validation from the training set (previously randomly sampled). The batch size was set to 8. As the optimizer, we chose *rmsprop*, because it is usually a good choice for RNNs and because most of our models are RNNs.

The metrics we use to compare the models are F1 score, as well as precision, recall, and accuracy. For the calculation of precision, recall, and the F1-score we used the *precision_recall_fscore_support* [3] method of scikit-learn [4].

CNN	BiLSTM (BOW)	BiLSTM	BiLSTM-CRF
0.81	0.80	0.72	0.73

Table 5.1: Comparison of the models using accuracy

The first metric we compare is accuracy. The CNN text classifier has the best result with an accuracy of 0.81, shown in Table 5.1. In addition, it was the model with the shortest runtime. The bi-LSTM model that uses the BOW model has the second best accuracy with 0.80. The bi-LSTM model that uses pre-trained word embeddings and the bi-LSTM model that uses a CRF layer on top had similar results regarding the accuracy, with 0.72 and 0.73. Among all the models the bi-LSTM with pre-trained word embeddings has the worst accuracy.

[3] sklearn metrics - https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html

[4] scikit learn - https://scikit-learn.org/stable/index.html

Accuracy gives us a first impression of how well the models perform. Because we do not work with a balanced dataset we need to further evaluate the models using the metrics precision, recall, and F1-score. The Table 5.2 summarizes our results regarding precision, recall, and F1-score.

Class	**CNN**			**BiLSTM (BOW)**			**BiLSTM**			**BiLSTM-CRF**		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
pi	0.88	0.85	0.83	0.89	0.85	0.87	0.90	0.82	0.86	0.96	0.78	0.86
bs	0.82	0.80	0.81	0.80	0.89	0.84	0.78	0.64	0.70	0.76	0.56	0.64
f	0.92	0.89	0.90	0.86	0.89	0.87	0.64	0.90	0.74	0.85	0.85	0.85
c	0.65	0.75	0.70	0.53	0.73	0.61	0.60	0.51	0.56	0.38	0.75	0.50
t	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
g	0.75	0.47	0.57	0.60	0.20	0.30	0.76	0.18	0.29	0.94	0.37	0.53
fi	0.84	0.97	0.90	0.86	0.98	0.92	0.74	0.99	0.85	0.77	0.99	0.87
o	0.59	0.46	0.52	0.52	0.11	0.18	0.67	0.01	0.03	0.53	0.13	0.21
Macro F1			0.65			0.58			0.50			0.50
Micro F1			0.81			0.80			0.72			0.73

Table 5.2: Comparison of the models using presicion, recall, and F1-score.
pi: personal information, bs: biographical sketch, f: family, c: characteristics, t: tribute, g: gratitude, fi: funeral information, o: other

The CNN text classifier has among all models the highest macro average F1-score with a value of 0.65, which can be seen in Table 5.2. This value results from the high values for the classes *family* and [funeral information], where both classes have a value of 0.90. It can be observed that the F1-score for the class *other* is higher compared to the other three models, with a value of 0.52. The reason lies in the recall value of 0.46. In comparison, the other models have a F1-score for the class *other* of less than 0.22, due to the low recall. The macro average F1 score for the bi-LSTM model is 0.58 which is a bit lower than for the CNN. The bi-LSTM has high F1-scores for the classes *personal information*, *family*, and *funeral information*, where each value except for the one of the class *family* are higher than the values for the CNN text classifier. The bi-LSTM model that uses the BOW model shows better result, regarding the F1 scores for each class, than the same model using word embeddings. This is also shown in the macro average F1-score of the bi-LSTM that is lower than the F1-score of the bi-LSTM using the BOW model, with a value of 0.50. We can observe that the F1 score is especially bad for the class *other* with a value of 0.03, due to a low recall of 0.01. The worst macro average F1-score among our models has the bi-LSTM (Word2Vec) and bi-LSTM-CRF with a value of 0.50. The F1-scores for the classes *other*, *gratitude*, and *family* are higher than those of the bi-LSTM using the word embeddings despite having a lover macro average F1-score. This shows that this model performs better for these classes than the previous.

Because of the little samples of sentences labeled as *tribute* none of our models predict a sentence as such, resulting in precision, recall, and F1 value of 0 for each model. The micro average F1-score for all models is exactly the same as the accuracy because our classification problem assigns for each instance one label. If we had allowed a multi-label classification, then the micro F1-score would be different.

5.3 Discussion

From the results presented in the previous section 5.2 we can deduce that the CNN model works best. Not only has the CNN model the highest accuracy with an accuracy of 0.81 but it has also the highest F1-score at 0.65. Apart from the high accuracy and F1-score it was also the only model that predicted the class *gratitude* as well as the class *other* better than the other models. The confusion matrix for the CNN model (Table 5.3) shows that it predicted the label *gratitude* 42 times and only falsely predicted a sentence labeled as gatitude as *funeral information* 32 times. The number of the falsely predicted classes is still higher in total than the correctly predicted class, but compared to the other models it made fewer mistakes regarding the class *gratitude*. Through the confusion tables of the different bi-LSTM models (Table 5.4-5.6) we can see that the models have difficulties with the prediction of the label *gratitude*. As with the CNN model, the other models falsely predict a sentence, which is actually labeled as *gratitude*, in most cases as *funeral information*. For the label *other* are the predictions of the CNN model the best. If one looks at the confusion matrix for the bi-LSTM with word embeddings (table **??**) and the bi-LSTM-crf (Table 5.6), one sees that the sentences with the label *other* are mostly classified as *funeral information*. For the bi-LSTM with the BOW model we can see that it falsely predicts the class *characteristics* instead of *other*.

The bi-LSTM model that uses a BOW model had a similar accuracy to the CNN model with an accuracy of 0.80 but a lower F1-score with a value of 0.58. If we take a look at the confusion matrix (Table 5.4) we can observe that in general the bi-LSTM model performed better for the classes *personal information*, *biographical sketch*, *characteristics*, and *funeral information*. The model performed worse for the classes *other* and *gratitude*, which is reflected in the F1-scores for the corresponding classes in Table 5.2. As we could previously observe in Section 5.2 the bi-LSTM using a BOW approach performs better than the other two models that are based on a bi-LSTM. This is also supported if we take a look at the confusion matrices Table 5.5 and Table 5.6. The bi-LSTM (BOW) predicts e.g., the class *biographical sketch* more correct than the other two. For this class, the model outperforms the other. Regarding the class *characteristics* has the bi-LSTM using word embeddings the lowest number for the true positive. Nevertheless, it performs in general slightly better than the bi-LSTM-CRF. Therefore we can say that despite similar accuracy and F1-score the bi-LSTM-CRF performs worse than the models without a CRF layer on top.

CNN Confusion Matrix

		Predicted							
		pi	bs	f	c	t	g	fi	o
Actual	pi	295	13	17	5	0	0	36	10
	bs	18	499	17	47	0	3	12	26
	f	6	16	675	16	0	5	10	33
	c	4	41	8	239	0	0	36	10
	t	0	2	2	0	0	0	0	0
	g	2	4	1	0	0	42	32	9
	fi	5	2	5	1	0	2	761	6
	o	6	33	11	62	0	4	49	143

Table 5.3: Confusion matrix of the CNN text classifier

Bi-LSTM (Word2Vec) Confusion Matrix

		Predicted							
		pi	bs	f	c	t	g	fi	o
Actual	pi	312	5	27	3	0	1	31	0
	bs	8	398	165	41	0	0	10	0
	f	16	7	683	9	0	0	45	0
	c	2	50	86	161	0	1	14	1
	t	0	2	2	0	0	0	0	0
	g	0	1	22	2	0	16	48	1
	fi	0	1	9	0	0	0	771	0
	o	10	45	80	49	0	3	117	4

Table 5.5: Confusion matrix of the bi-LSTM (Word2Vec) sequence labeler

Bi-LSTM (BOW) Confusion Matrix

		Predicted							
		pi	bs	f	c	t	g	fi	o
Actual	pi	324	24	12	3	0	0	15	1
	bs	2	552	29	31	0	2	3	3
	f	9	22	677	17	0	5	21	14
	c	8	53	21	229	0	1	3	0
	t	0	1	2	0	0	0	0	1
	g	0	5	9	19	0	18	32	7
	fi	2	0	5	3	0	0	765	6
	o	20	33	35	128	0	9	49	34

Table 5.4: Confusion matrix of the bi-LSTM (BOW) sequence labeler

Bi-LSTM-CRF Confusion Matrix

		Predicted							
		pi	bs	f	c	t	g	fi	o
Actual	pi	295	20	22	9	0	0	31	2
	bs	0	347	29	231	0	1	9	5
	f	4	27	644	30	0	0	39	16
	c	1	33	29	236	0	0	8	8
	t	0	1	1	2	0	0	0	0
	g	0	3	5	8	0	33	39	2
	fi	0	2	4	0	0	0	772	3
	o	6	23	23	110	0	1	104	41

Table 5.6: Confusion matrix of the bi-LSTM-CRF sequence labeler

Notes: pi: personal information, bs: biographical sketch, f: family, c: characteristics, t: tribute, g: gratitude, fi: funeral information, o: other

6 Conclusion

This work dealt with the question if we can automatically structure obituaries. Therefore, we presented a new corpus consisting of 20058 obituaries from which we annotated 1008. We implemented and tested four models, a CNN text classifier, a bi-LSTM network using a BOW model and one using word embeddings, and a bi-LSTM-CRF on our annotated dataset. We compared them based on accuracy, precision, recall, and F1-score.

From our results we concluded that the CNN text classifier produced the best results with an accuracy of 0.81, considering the experimental settings, and the highest macro average F1-score of 0.65. The bi-LSTM (BOW) model generated also good results and even better regarding the classes *personal information* and *biographical sketch*, which makes it also a viable model for our dataset.

Our work enables future research, showing that the structuring of obituaries is viable, which can be used for data mining. Through zoning, it will be possible to address questions such as the if there is a correlation between work and cause of death, are where cultural differences between obituaries from different countries, or we can examine obituaries for gender stereotypes.

References

[1] M. Alfano, A. Higgins, J. Levernier. “Identifying virtues and values through obituary data-mining”. In: *The Journal of Value Inquiry* 52.1 (2018), pp. 59–79 (cit. on pp. 19, 20).

[2] C. Bai. *Best Obituary Examples and Free Templates, Format for Newspaper*. 2017. URL: https://sympathies.co/best-obituary-examples-templates/ (cit. on pp. 13, 21, 22).

[3] D. Bamman, N. A. Smith. “Unsupervised discovery of biographical structure from text”. In: *Transactions of the Association for Computational Linguistics* 2 (2014), pp. 363–376 (cit. on p. 20).

[4] E. Chen. *Introduction to Conditional Random Fields*. 2012. URL: http://blog.echen.me/2012/01/03/introduction-to-conditional-random-fields/ (cit. on p. 17).

[5] C. Epstein, R. Epstein. “Death in The New York Times: the price of fame is a faster flame”. In: *QJM: An International Journal of Medicine* 106.6 (2013), pp. 517–521 (cit. on p. 19).

[6] X. Feng, Y. Liang, X. Shi, D. Xu, X. Wang, R. Guan. “Overfitting Reduction of Text Classification Based on AdaBELM”. In: *Entropy* 19.7 (2017), p. 330 (cit. on p. 15).

[7] C. W. Ford, C.-C. Chiang, H. Wu, R. R. Chilka, J. R. Talburt. “Text data mining: a case study”. In: *Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on*. Vol. 1. IEEE. 2005, pp. 122–127 (cit. on p. 19).

[8] Y. Guo, A. Korhonen, T. Poibeau. “A weakly-supervised approach to argumentative zoning of scientific documents”. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. 2011, pp. 273–283 (cit. on p. 18).

[9] K.-S. Han. “Personal Information Extraction from Korean Obituaries”. In: *IEICE TRANSACTIONS on Information and Systems* 96.12 (2013), pp. 2873–2876 (cit. on p. 20).

[10] M. Herat. “Avoiding the reaper: Notions of death in Sri Lankan obituaries”. In: *International Journal of Language Studies* 8.3 (2014), pp. 117–144 (cit. on p. 19).

[11] Z. Huang, W. Xu, K. Yu. “Bidirectional LSTM-CRF models for sequence tagging”. In: *arXiv preprint arXiv:1508.01991* (2015) (cit. on p. 17).

[12] E. Ikonomakis, S. Kotsiantis, V. Tampakas. “Text Classification Using Machine Learning Techniques”. In: *WSEAS transactions on computers* 4 (Aug. 2005), pp. 966–974 (cit. on p. 13).

[13] T. Joachims. “Transductive inference for text classification using support vector machines”. In: *ICML*. Vol. 99. 1999, pp. 200–209 (cit. on pp. 13, 15).

[14] R. Johnson, T. Zhang. “Effective Use of Word Order for Text Categorization with Convolutional Neural Networks”. In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Denver, Colorado: Association for Computational Linguistics, May 2015, pp. 103–112. DOI: 10.3115/v1/N15-1011. URL: https://www.aclweb.org/anthology/N15-1011 (cit. on pp. 15, 16).

[15] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov. “Bag of tricks for efficient text classification”. In: *arXiv preprint arXiv:1607.01759* (2016) (cit. on p. 13).

[16] G. Paaß, I. Konya. “Machine learning for document structure recognition”. In: *Modeling, Learning, and Processing of Text Technological Data Structures*. Springer, 2011, pp. 221–247 (cit. on p. 18).

[17] B. Settles, M. Craven. “An analysis of active learning strategies for sequence labeling tasks”. In: *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics. 2008, pp. 1070–1079 (cit. on p. 17).

[18] F. Sha, F. Pereira. “Shallow parsing with conditional random fields”. In: *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics. 2003, pp. 134–141 (cit. on p. 17).

[19] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, M. Demirbas. “Short text classification in twitter to improve information filtering”. MA thesis. The Ohio State University, 2010 (cit. on pp. 13, 15).

[20] K. Tomanek, U. Hahn. “Semi-supervised active learning for sequence labeling”. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*. Association for Computational Linguistics. 2009, pp. 1039–1047 (cit. on pp. 13, 16, 17).

[21] A. K. Uysal. “An improved global feature selection scheme for text classification”. In: *Expert systems with Applications* 43 (2016), pp. 82–92 (cit. on pp. 13, 15).

[22] T. Wang, D. J. Wu, A. Coates, A. Y. Ng. “End-to-end text recognition with convolutional neural networks”. In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. Nov. 2012, pp. 3304–3308 (cit. on p. 16).

[23] Z. Xing, J. Pei, E. Keogh. “A brief survey on sequence classification”. In: *ACM Sigkdd Explorations Newsletter* 12.1 (2010), pp. 40–48 (cit. on pp. 13, 16).

[24] L. Xu, D. W. Embley. “Categorisation of web documents using extraction ontologies”. In: *International Journal of Metadata, Semantics and Ontologies* 3.1 (2008), pp. 3–20 (cit. on p. 20).

All links were last followed on May 13, 2019.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

place, date, signature