Automatic summarization of medical conversations, a review

Jessica López Espejel [1, 2]
(1) CEA, LIST, DIASI, F-91191 Gif-sur-Yvette, France.
(2) Paris 13 University, LIPN, 93430 Villateneuse, France.
jessica.lopez-espejel@cea.fr

RÉSUMÉ

L'analyse de la conversation joue un rôle important dans le développement d'appareils de simulation pour la formation des professionnels de la santé (médecins, infirmières). Notre objectif est de développer une méthode de synthèse automatique originale pour les conversations médicales entre un patient et un professionnel de la santé, basée sur les avancées récentes en matière de synthèse à l'aide de réseaux de neurones convolutionnels et récurrents. La méthode proposée doit être adaptée aux problèmes spécifiques liés à la synthèse des dialogues. Cet article présente une revue des différentes méthodes pour les résumés par extraction et par abstraction et pour l'analyse du dialogue. Nous décrivons aussi les utilisation du Traitement Automatique des Langues dans le domaine médical.

ABSTRACT

Conversational analysis plays an important role in the development of simulation devices for the training of health professionals (doctors, nurses). Our goal is to develop an original automatic synthesis method for medical conversations between a patient and a healthcare professional, based on recent advances in summarization using convolutional and recurrent neural networks. The proposed method must be adapted to the specific problems related to the synthesis of dialogues. This article presents a review of the different methods for extractive and abstractive summarization, and for dialogue analysis. We also describe the use of Natural Language Processing in the medical field.

MOTS-CLÉS : Résumé automatique, domaine médical, dialogues, revue.

KEYWORDS: Automatic summarization, medical domain, dialogues, review.

1 Introduction

Spoken medical dialogues are common and useful among doctors, nurses and patients. Automatically gathering the information of these dialogues is relevant for the medical practitioners and patients. For instance, the doctor can generate a medical history without omitting information, and a patient can review the diagnostic and the treatment later in time. However, none of them want a full record or transcript all dialogues. Besides, doctors and nurses have many interactions with different patients all day long, so it wouldn't be productive that they now have to deal with all these records. Accordingly, a challenge for Natural Language Processing (NLP) arises : obtaining automatic summaries from medical conversations.

According to the above, our final goal is to develop an automatic summarization method for medical conversations between patients and doctors or nurses. However, our work is related to a project in our laboratory implementing a system for conversations between a virtual patient (chatbot) and medicine students (Laleye *et al.*, 2019). In this serious game, the student establishes a conversation with the chatbot to establish its diagnostic about the medical condition of the patient. The proposed method should be adapted to the specific problems of summarizing dialogues based on recent methods used in automatic summarization, such as recurrent neural networks. The hypothesis is that pertinent segments of dialogue might be detected and abstracted for their inclusion in a summary through deep learning. We suppose that, by identifying pertinent blocks of dialogues, it would be easier to detect important clinical issues and discard issues fragments like greetings, acknowledgments, social concerns, etc.

In this paper we review the work related to our goal of summarizing medical conversations. We first describe the two main types of summaries : extractive and abstractive. Extractive methods consist of copying chunks, usually sentences, from the original text. Abstractive methods are able to generate new words and sentences. In the literature there are more extractive approaches since they were easier than abstractive methods, even if the latter are becoming conceivable (See *et al.*, 2017). This paper describes different techniques for each one. Besides, it also reviews the specific mechanisms used in research on dialogue summarization.

The last section is crucial because it describes different approaches and works in NLP in the medical domain. These approaches include medical conversation systems and summarization in the medical domain. We discuss the challenges they represent.

2 Methods for summarization

In this section we will mention different types of approaches that researchers have been working along the time to get automatic summaries.

2.1 Summarization criteria

In the following lines we will describe two important criteria to summarize text. The first criteria is based on the frequency of the words. The second criteria are focused on the text features to detect the importance of the sentences.

Frequency criteria Over the years, several criteria have been developed to generate extractive summaries. One of the most cited in the literature is TF-IDF. TF (Term-Frequency) was proposed by Luhn (1958) and is the frequency of a word in the document. IDF (Inverse Document Frequency) was proposed by Sparck Jones (1972) and attenuates the weight of words that appear in a lot of the documents of the collection and increases the weight of words that occur in a few of them. The first works in summarization were based on $TF - IDF$. For instance, Wang & Cardie (2011) used unsupervised methods like TF-IDF, LDA, topic modeling and supervised clustering to produce a concise abstract of the decisions taken in spoken meetings.

Surface features criteria An alternative to detect the relevance of a sentence is through features of different kind. Yogan *et al.* (2016) mentioned the following features : title/headline words, sentence position, sentence length, proper noun and term weight. Chopra *et al.* (2016) captured the word position in which it occurs in the sentence and its context in which appears in the sentence. Lacson *et al.* (2006) used the length of dialogue turns to detect important information in a conversation. Nallapati *et al.* (2016) combined features to get better results.

2.2 Summarization methods

Probabilistic Models There are also probabilistic models, such as Context Free Grammars and Markov Models (MM). Probabilistic Context Free Grammars (PCFG) is a probabilistic model of syntax for tree structures. Rahman *et al.* (2001) worked on automatic summarization of Web pages and used PCFG to define syntactic structures, analyze and understand content. Besides, Knight & Marcu (2002) worked on sentence compression and developed a probabilistic noisy-channel model that used PCFG to assign probabilities to a tree. On the other hand, Chen & Withgott (1992) used Markov Model(MM) on speech summarizing. Jing & McKeown (1999) proposed an algorithm based on Hidden Markov Model(HMM) that decomposes human-written summary sentences to determine the relations between the sentences in a summary written by humans and sentences in the original text. Conroy & O'leary (2001) proposed a method for text summarization, which considers three features : the position of the sentence in the document (using Hidden Markov Model), the number of terms in the sentence and the probabilities of the terms.

Optimization methods On the other hand, other important approach to get summaries from text is based on Integer Linear Program (ILP). Gillick & Favre (2009) used ILP to exact inference under a maximum coverage model in automatic summarization and sentence compression. Mnasri *et al.* (2017) used the same methods for multidocument update summarization, improving it by taking into semantic similarity and document structure.

Graph-based methods Mihalcea (2004) used graph-based ranking algorithms to extract the most important sentences from DUC (2002). Unlike Mihalcea (2004), Litvak & Last (2008) used graphs to identify keywords to be used in extractive summarization of text documents.

Machine Learning Approaches In recent years, researchers have proposed methods based-on machine learning to summarize text. Naive Bayes was used by Kupiec *et al.* (1995) to chose if a sentence belongs or not to a summary. Recently Ramanujam & Kaliappan (2016) extended the application of the Naive Bayes algorithm to automatic summarization in multi-documents. Aside from Nayve Bayes, clustering algorithms have been used by Aliguliyev (2009) and KM & Soumya (2015) to get extractive summaries. Aliguliyev (2009) proposed a method based on sentence clustering, while KM & Soumya (2015) prepared the cluster center using a n-dimensional vector space and the documents similarity is measure by cosine similarity to generate the documents clusters.

Besides, researchers have used Support Vector Machine (SVM). For example, (Schilder & Kondadadi, 2008) work on query-based multi-document summarization and they use SVM to rank all sentences in the topic cluster for summarization. Another algorithm that researchers have been used is genetic algorithms. For instance, Chatterjee *et al.* (2012) worked on extractive summaries representing the document as a Direct Acyclic grapth (DAC). They used genetic algorithm to maximize the fitness function and get the summary. After few years, Bossard & Rodrigues (2017) used objective function from generic algorithms to explore summaries space and compute the probability distribution of tokens.

One of the most widely used family of machine learning methods for automatic summarization is Neural Networks (NN). In NLP, the currently most relevant recurrent neural networks are : LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit). LSTM has the ability to add or remove information through structures called gates : forget gate, input gate, candidate gate and output gate layer. Meanwhile, GRU is a variant of LSTM and it combines the forget and input gates into a single *update gate*. LSTM and GRU present some advantages, such as the ability to store long-term dependencies, they avoid the problem of vanishing gradient and they consider the order of the words. In abstractive text summarization, taking into account the order of words is one of the greatest advances because the summaries present more coherence.

Kaikhah (2004) used NN to learn relevant features of sentences and decide if these sentences should be included in the summary, while Sinha *et al.* (2018) proposed a model based on feed-forward networks for single document summarization. LSTM was used by Cheng & Lapata (2016) and Zhou (2016). Both used an encoder-decoder model. However Cheng & Lapata (2016) worked on extractive summarization of single documents, while Zhou (2016) applied a hierarchical LSTM model for building the sentence representations in abstractive and long summaries. GRUs were used by Nallapati *et al.* (2017) and more recently by Li *et al.* (2018). Nallapati *et al.* (2017) proposed SummaRuNNer (simple recurrent network based sequence classifier) for extractive summarization. Li *et al.* (2018) extended the basic encoder-decoder model. They added an intermediate layer (select layer) that consists in two parts : gated global information filtering and local sentence selection.

Over the past few years, the number of works in abstractive automatic summarization has increased. One of the first NN approaches to tackle abstractive summarization was the Attention RNN (Recurrent NN). Encoder-Decoder Model which was introduced by Bahdanau *et al.* (2015). Rush *et al.* (2015) combine this approach with a generation algorithm for a fully data-driven approach to abstractive sentence summarization. One extension of this work for the same problem was developed by Chopra *et al.* (2016) but they didn't use feed-forward neural language model for generation like Rush *et al.* (2015). They used a convolutional attention-based conditional recurrent neural network model, producing further improvement in performance on Gigaword and DUC datasets.

The previous mentioned works focused on short texts. However Nallapati *et al.* (2016) proposed a new dataset for the task of abstractive summarization of a document into multiples sentences and establish brenchmarks. Furthermore, See *et al.* (2017) used this dataset and proposed a hybrid architecture between sequence-to-sequence attention model and a pointer network Vinyals *et al.* (2015) that facilitates copying words from the source text via pointing or generating words from a fixed vocabulary. They also adapt the coverage model of Tu *et al.* (2016) to solve the problem of repetitions in the generated summary.

Automatic summarization methods have been applied on various kinds of documents, such as text (news, articles, etc), dialogues, in medical and other domains. In the literature, we can find automatic summarization of dialogues in meetings, recorded conversations in call centers and other events that happen everyday. The majority of works on medical summarization have focused on research articles. However in our case we are interested to get automatic summaries from medical conversations. In sections 3 and 4, we describe independently NLP works on dialogue analysis and the medical domain, with a focus on summarization.

3 Summarization methods applied on dialogues

A dialogue is a sequence of conversational turns between multiple participants, where each turn would modify each participant cognitive status and the current dialogue state Chih-Wen & Chen (2018). However, we suppose that two people are enough to establish a dialogue. Currently, a person can also establish a conversation with a system equipped with conversational intelligence Turing (1950), either in written or oral form.

The process of automatic dialogue summarization is a challenging summarization task because, e.g., unlike in journalistic texts, the most important sentence is not the first one in each paragraph, and we have to consider other aspects such as the speakers' turns. In the case of capturing the conversations orally, we face several types of disfluencies such as fillers, repetitions, repairs, or unfinished clauses Zechner & Waibel (2000). Otherwise, on conversations obtained from written sources, it might be necessary to pre-process the data due to lexical irregularities.

Over the past few years, researchers have worked on automatic summaries in different areas of daily life, such as mettings, phone calls and medical domain. For example, Hendrik Buist *et al.* (2004) worked on audio-visual meetings and their goal was to extract all important topics. In another related work, the goal was to produce a concise abstract of decisions taken in spoken meetings Wang & Cardie (2011).

On the other hand, automatic summary systems dedicated to call centers have also been developed to record calls between agents and clients, to know the content of the call and the agent's level of expertise. Tamura *et al.* (2011) proposed an extractive method and they introduced a component that removes frequent sentences from summary, in comparison with (Evgeny *et al.*, 2015) where they focus on abstractive summarization and used domain knowledge to fill hand-written templates from entities detected in the transcript of the conversation.

Additionally, the need to use systems of dialogue in the medical domain has increased in the last two decades Bickmore & Giorgino (2006) . The HOMEY system Piazza *et al.* (2004) was developed to enhance communication between health center specialists and patients with chronic diseases. After few years, Andrenucci (2008) used the automatic question-answering (QA) paradigm in medical domain. They determined that the best approaches for medical applications are deep analysis of language (semantics) and template based.

4 NLP in Medical Domain

In medical domain, doctors and nurses have a lot of information about patients : medical records, appointments, schedule of activities, etc. All information must be organized, and manually involves a lot of time and human effort. Besides, the amount of information increases each second and there is sometimes not enough people.

Currently, there are algorithms capable of helping nurses and doctors to summarize and organize information. However the medical domain is huge. Some works are focused on medical diagnostic tools, such as Doan *et al.* (2016). They detected child patients with high suspicion of Kawasaki Disease (KD) based on standard clinical terms and medical lexicon usage. Likewise, researchers have worked on the detection of depression, suicide risk Pestian *et al.* (2010) Mulholland & Quinn (2013) and mental diseases Thomas *et al.* (2005) Karlekar *et al.* (2018) by clinical notes and social media

Calvo *et al.* (2017) analysis using NLP techniques.

Other researches have developed intelligent agents, such as Allen *et al.* (2006) to help people at home. People can talk to the systems using natural language.

The following subsections concentrate on conversational and summarization systems in the medical domain.

4.1 Medical Conversational Systems

There are many conversational applications that have been developed to help the health field, such as intelligence agents and health dialog systems. Allen *et al.* (2006) developed Chester, a prototype of medical advisor. Chester provides information, advises based on prescribed medications and reports back to medical support team, it can answer questions asked by users. In the meantime, de Rosis *et al.* (2006) developed a conversational agent to talk to patients to influence them to change their dietary behavior.

Migneault *et al.* (2006) describe the approach of how to write dialogues for TLC (Telephone-Linked Care) telephone systems, whose objective is to offer health-related services. At the same time, (Beveridge & Fox, 2006) studied the automatic generation of dialogue combining knowledge of the structure of tasks and ontological knowledge, the objective is to decide if a patient must be referred or not with a cancer specialist.

4.2 Summarization in medical domain

Automatic summarization can be used for medical consultations. The dialogue between patients and doctors can be recorded and a summary been generated with the most important points, to the attention of other doctors and for the medical history of the patient. This is the consultation report.

To the best of our knowledge, there are few works about summarization on medical dialogues. One of them was developed by Lacson *et al.* (2006). They worked on automatic summarization of dialogues between nurses and dialysis patients. Their system consists in two main components : induction (machine-learning algorithm for classifying dialogue turns) and summarization (prediction based on a meaning representation encoded in lexical and contextual features). Unlike them, Sarkar *et al.* (2011) worked on automatic summarization of medical articles to decide if an article is useful or not. The algorithm classify the sentences as positive or negative. After ranking, sentences are selected to generate the final summary.

In addition, researchers have also developed NLP tools focused on medical domain that can help us to generate better quality of summaries through the identification of medical terms. For instance, Li & Wu (2006) implemented KIP (Keyphrase Identification Program) to identify topical concepts from medical documents. KIP combines noun phrase extraction and keyphrase identification. MedPost is a Part-Of-Speech(POS) tagger developed by Smith *et al.* (2004). MedPost is based on Hidden Markov Model. Additionally, Tanabe & Wilbur (2002) worked on extracting gene and proteins names from MEDLINE documents.

4.3 Challenges on medical domain

Clinical notes and medical texts present several challenges for NLP. One of the most important challenge is based on word disambiguation because extracting the meaning from unstructured text is not easy. For example, *cold* can refer to the weather or a disease Townsend (2013).

Besides, clinical texts are often ungrammatical. It means that they contain incomplete sentences and limited context. Also, along the time new terms and abbreviations emerge. A lot of this terms that appear in medical sources such as the Unified Medical Language System (UMLS) Metathesaurus have multiple meanings which depend on the context.

Otherwise, an important challenge on summarization area is to get summaries from medical dialogues. Even finding or building corpus of medical conversations is hard.

Finally, an important challenge in the medical domain is ethical. The building of systems must ensure that private sensitive data cannot leak. Nobody except authorized medical people must be able to access data related to a specific patient. This poses hard to solve problems of anonymisation. Besides, it is very important not to modify the meaning or to give incomplete information in the final summary.

5 Discussion

We have presented works about medical summarization divided in three sections : methods for summarization, NLP for dialogue analysis and medical conversation systems. In the first section, the most used methods on extractive and abstractive summaries have been reviewed. The second section presented works focused on dialogue summaries, and the challenge that they present. In the last section we cited works to help people at home, and software to detect depression, suicide and mental diseases.

After this review of the bibliography, we realized that there are a lot of articles about automatic summarization, but the majority are focused on scientific articles, news, etc. and there are few works on conversations. Furthermore, there is even less works on medical dialogue. One of the most important reason is that it is not easy to get datasets. Similarly, we find more articles on extractive summaries than abstractive summaries, even if the current tendency is clearly to switch to abstractive using neural networks.

Our future work is based on the model by See *et al.* (2017). We will apply this technique on the PubMed dataset (medical articles). We decided to work with this algorithm because they used a hybrid pointer-generator network to decide if a specific word is copied from the source text or if a new word from the vocabulary will be generated. Our hypothesis is that pointer-generator networks can be useful on the medical domain because we can generate new words from its specialized vocabulary. This will make easier to handle medical terms. Besides, they proposed a novel variant of coverage based on Tu *et al.* (2016) to avoid repetitions in summaries. As noted above, it will be very important to insure that the generated text doesn't modify the sense of the original dialogue. We will also make the algorithm able to handle long texts while it was designed only for the first words of a news article.

Our second approach is to work on the AMI corpus developed by Mccowan *et al.* (2005), a dataset of 100 hours of multi-modal meetings recordings. We will develop a mechanism to generate abstractive summaries from dialogues. We decided to work on AMI dataset to get automatic summaries from real conversations and consider aspects that we can find only in dialogues such as greetings, speakers' turns, etc.

However, our main idea is to develop a mechanism to be able to get abstractive summaries from medical dialogues. To achieve this task, it's necessary to work with both approaches : summarization of dialogues and summarization in the medical domain.

Références

ALIGULIYEV R. M. (2009). A new sentence similarity measure and sentence based extractive technique for automatic text summarization. *Expert Systems with Applications*, **36(4)**, 7764–7772.

ALLEN J. F., FERGUSON G., BLAYLOCK N., BYRON D. K., CHAMBERS N., DZIKOVSKA M. O., GALESCU L. & SWIFT M. D. (2006). Chester : Towards a personal medication advisor. *Journal of Biomedical Informatics*, **39(5)**, 500–513.

ANDRENUCCI A. (2008). Automated question-answering techniques and the medical domain. In *HEALTHINF*.

BAHDANAU D., CHO K. & BENGIO Y. (2015). Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

BEVERIDGE M. & FOX J. (2006). Automatic generation of spoken dialogue from medical plans and ontologies. *Journal of Biomedical Informatics*, **39**(5), 482–499.

BICKMORE T. W. & GIORGINO T. (2006). Health dialog systems for patients and consumers. *Journal of biomedical informatics*, **39(5)**, 556–71.

BOSSARD A. & RODRIGUES C. (2017). An evolutionary algorithm for automatic summarization. In *Proceedings of Recent Advances in Natural Language Processings 2017*.

CALVO R., MILNE D. N., HUSSAIN M. S. & CHRISTENSEN H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, **23**, 649–685.

CHATTERJEE N., MITTAL A. & GOYAL S. (2012). Single document extractive text summarization using genetic algorithms. In *Third International Conference on Emerging Applications of Information Technology*, p. 19–23.

CHEN F. & WITHGOTT M. (1992). The use of emphasis to automatically summarize a spoken discourse. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, p. 229–232.

CHENG J. & LAPATA M. (2016). Neural summarization by extracting sentences and words. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, p. 484–494.

CHIH-WEN G. & CHEN Y.-N. (2018). Abstractive dialogue summarization with sentence-gated modeling optimized by dialogue acts. *CoRR*.

CHOPRA S., AULI M. & RUSH A. M. (2016). Abstractive sentence summarization with attentive recurrent neural networks. In *NAACL-HLT*, p. 93–98.

CONROY J. M. & O'LEARY D. P. (2001). Text summarization via hidden markov models. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, p. 406–407.

DE ROSIS F., NOVIELLI N., CAROFIGLIO V., CAVALLUZZI A. & DE CAROLIS B. (2006). User modeling and adaptation in health promotion dialogs with an animated character. *Journal of Biomedical Informatics*, **39(5)**, 514–531.

DOAN S., MAEHARA C. K., CHAPARRO J., LU S., LIU R., GRAHAM A., BERRY E., HSU C.-N., T. KANEGAYE J., D. LLOYD D., OHNO-MACHADO L., C. BURNS J. & TREMOULET A. (2016). Building a natural language processing tool to identify patients with high clinical suspicion for kawasaki disease from emergency department notes. *Academic Emergency Medicine*, **23**, 328–636.

DUC (2002). Documentunderstandingconference2002. urlhttps ://www-nlpir.nist.gov/projects/duc/.

EVGENY S., FAVRE B., FIROJ A., SHAMMUR C., SINGLA K., TRIONE J., B'ECHET F. & GIUSEPPE R. (2015). Automatic summarization of call-center conversations. In *Conference on Empirical Methods in Natural Language Processing*.

GILLICK D. & FAVRE B. (2009). A scalable global model for summarization. In *Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing*, p. 10–18 : Association for Computational Linguistics.

HENDRIK BUIST A., KRAAIJ W. & RAAIJMAKER S. (2004). Automatic summarization of meeting data : A feasibility study. In *Proc. Meeting of Computational Linguistics in the Netherlands (CLIN)*.

JING H. & MCKEOWN K. R. (1999). The decomposition of human-written summary sentences. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, p. 129–136 : ACM.

KAIKHAH K. (2004). *Text Summarization using Neural Networks*. PhD thesis, Department of Computer Science, Texas State university.

KARLEKAR S., NIU T. & BANSAL M. (2018). Detecting linguistic characteristics of alzheimer's dementia by interpreting neural models. In *NAACL-HLT*, volume 2, p. 701–707.

KM S. & SOUMYA R. (2015). Text summarization using clustering technique and svm technique. *International Journal of Applied Engineering Research*, **10**, 25511–25519.

KNIGHT K. & MARCU D. (2002). Summarization beyond sentence extraction : A probabilistic approach to sentence compression. *Artificial Intelligence*, **139**, 91–107.

KUPIEC J., PEDERSEN J. & CHEN F. (1995). A trainable document summarizer. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, p. 68–73.

LACSON R. C., BARZILAY R. & LONG W. J. (2006). Automatic analysis of medical dialogue in the home hemodialysis domain : Structure induction and summarization. *Journal of Biomedical Informatics*, **39(5)**, 541–555.

LALEYE F. A. A., BLANIÉ A., BROUQUET A., BENHAMOU D. & DE CHALENDAR G. (2019). Hybridation d'un agent conversationnel avec des plongements lexicaux pour la formation au diagnostic médical. In *23ème Conférence Sur Le Traitement Automatique Des Langues Naturelles (TALN-RECITAL 2019)*, Toulouse, France.

LI Q. & WU Y.-F. (2006). Identifying important concepts from medical documents. *Journal of biomedical informatics*, **39**, 668–679.

LI W., XIAO X., LYU Y. & WANG Y. (2018). Improving neural abstractive document summarization with explicit information selection modeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

LITVAK M. & LAST M. (2008). Graph-based keyword extraction for single-document summarization. In *Proceedings of the Workshop on Multi-source Multilingual Information*, p. 17–24.

LUHN H. (1958). The automatic creation of literature abstracts. *IBM Journal of Research Development*, **2**, 159–165.

MCCOWAN I., LATHOUD G., LINCOLN M., LISOWSKA A., POST W., REIDSMA D. & WELLNER P. (2005). The ami meeting corpus. In *5th International Conference on Methods and Techniques in Behavioral Research.*

MIGNEAULT J. P., FARZANFAR R., WRIGHT J. A. & FRIEDMAN R. H. (2006). How to write health dialog for a talking computer. *Journal of Biomedical Informatics*, **39**, 468–481.

MIHALCEA R. (2004). Graph-based ranking algorithms for sentence extraction, applied to text summarization. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, p. 170–173.

MNASRI M., DE CHALENDAR G. & FERRET O. (2017). Taking into account inter-sentence similarity for update summarization. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2 : Short Papers)*, p. 204–209, Taipei, Taiwan : Asian Federation of Natural Language Processing.

MULHOLLAND M. & QUINN J. (2013). Suicidal tendencies : The automatic classification of suicidal and non-suicidal lyricists using nlp. In *6th International Joint Conference on Natural Language Processing*, p. 19–28.

NALLAPATI R., ZHAI F. & ZHOU B. (2017). Summarunner : A recurrent neural network based sequence model for extractive summarization of documents. In *AAAI*.

NALLAPATI R., ZHOU B., NOGUEIRA DOS SANTOS C., GÜLÇEHRE C. & XIANG B. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, p. 280–290.

PESTIAN J. P., NASRALLAH H. A., MATYKIEWICZ P., BENNETT A. & LEENAARS A. (2010). Suicide note classification using natural language processing : A content analysis. *Biomedical informatics insights*, **2010 (3)**, 19–28.

PIAZZA M., GIORGINO T., AZZINI I., STEFANELLI M. & LUO R. (2004). Cognitive human factors for telemedicine systems. *Studies in Health Technology and Informatics*, p. 974–978.

RAHMAN A., ALAM H., HARTONO R. & ARIYOSHI K. (2001). Automatic summarization of web content to smaller display devices. In *ICDAR Proceedings of Sixth International Conference on Document Analysis and Recognition*, p. 1064–1068.

RAMANUJAM N. & KALIAPPAN M. (2016). An automatic multidocument text summarization approach based on naïve bayesian classifier using timestamp strategy. In *The Scientific World Journal*.

RUSH A. M., CHOPRA S. & WESTON J. (2015). A neural attention model for sentence summarization. *Association for Computational Linguistics*, **Conference on Empirical Methods in Natural Language Processing**, 379–389.

SARKAR K., NASIPURI M. & GHOSE S. (2011). Using machine learning for medical document summarization. *International Journal of Database Theory and Application*, **4**.

SCHILDER F. & KONDADADI R. (2008). Fastsum : Fast and accurate query-based multi-document summarization. In *Association for Computational Linguistics*, p. 205–208.

SEE A., LIU P. J. & MANNING C. D. (2017). Get to the point : Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, p. 1073–1083.

SINHA A., YADAV A. & GAHLOT A. (2018). Extractive text summarization using neural networks. *CoRR*, **abs/1802.10137**.

SMITH L., RINDFLESCH T. & WILBUR W. (2004). Medpost : A part-of-speech tagger for biomedical text. *Bioinformatics (Oxford, England)*, **20**, 2320–1.

SPARCK JONES K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, **28**, 11–21.

TAMURA A., ISHIKAWA K., SAIKOU M. & TSUCHIDA M. (2011). Extractive summarization method for contact center dialogues based on call logs. *5th International Joint Conference on Natural Language Processing*, p. 8–13.

TANABE L. & WILBUR W. J. (2002). Tagging gene and protein names in full text articles. In *Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain*, volume 20, p. 9–13.

THOMAS C., KESELJ V., CERCONE N., ROCKWOOD K. & ASP E. (2005). Automatic detection and rating of dementia of alzheimer type through lexical analysis of spontaneous speech. *IEEE International Conference Mechatronics and Automation, 2005*, **3**, 1569–1574.

TOWNSEND H. (2013). Natural language processing and clinical outcomes : the promise and progress of nlp for improved care. *Journal of AHIMA / American Health Information Management Association*, **84**, 44–5.

TU Z., LU Z., LIU Y., LIU X. & LI H. (2016). Modeling coverage for neural machine translation. In *ACL*.

TURING A. M. (1950). Computing machinery and intelligence. *Mind*, **49**, 433–460.

VINYALS O., FORTUNATO M. & JAITLY N. (2015). Pointer networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*.

WANG L. & CARDIE C. (2011). Summarizing decisions in spoken meetings. In *Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages*, p. 16–24.

YOGAN J. K., GOH O. S., HALIZAH B., NGO H. C. & PUSPALATA C. S. (2016). A review on automatic text summarization approaches. **12**, 178–190.

ZECHNER K. & WAIBEL A. H. (2000). Diasumm : Flexible summarization of spontaneous dialogues in unrestricted domains. In *International Conference on Computational Linguistics*.

ZHOU Z. (2016). A hierarchical model for text autosummarization. `https://cs224d.stanford.edu/reports/zhenpeng.pdf`.