

Towards Automatic Generation of Context-Based Abstractive Discharge Summaries for Supporting Transition of Care

Diana Diaz^{1*}, Celia Cintas², William Ogallo² and Aisha Walcott-Bryant²

¹Department of Computer Science, Wayne State University, Detroit, MI 48202

²IBM Research Africa, Nairobi, Kenya

dmd@wayne.edu, celia.cintas@ibm.com, william.ogallo@ibm.com, awalcott@ke.ibm.com

Abstract

Discharge summaries are essential for the transition of patients' care but often lack sufficient information. We present an attention-based model to generate discharge summaries to support communication during the transition of care from intensive care units (ICU) to community care. We trained and evaluated our approach over 500,000 clinical progress notes. The summaries automatically generated by our model achieve a ROUGE-L of 0.83 when comparing with discharge summaries written by health professionals. We attribute the high performance to our three-step pipeline that incorporates disease and specialist contexts to enrich the summaries with relevant information based on the context of the hospital stay. Additionally, we present a novel visualization of ICU flow of care using MIMIC-III. Our promising results have the potential to improve the pipeline of hospital discharge and continuous health care.

1 Introduction

Patients who transition from intensive care to regular care are considered at high risk for adverse events, such as an unintended injury or accident that results in disability or readmission to ICU [Woloshynowych *et al.*, 2003; Chaboyer *et al.*, 2008]. Discharge summaries are the most common means of communication between inpatient and outpatient providers. However, numerous studies have shown that discharge summaries often fail to provide important information, such as the primary diagnosis, results of abnormal diagnostics, and details about the hospital course [Kripalani *et al.*, 2007b; Kripalani *et al.*, 2007a]. This could be due to the well-known problem of overworked physicians. In the United States for example, during a working day, physicians spend only 27% of their time on direct clinical face time with patients and 49.2% of their time on data entry in Electronic Health Records (EHRs) and desk work [Sinsky *et al.*, 2017]. Additionally, up to 54% of physicians experienced some sign of burnout in 2014 [Sinsky *et al.*, 2017].

Consequently, reducing time spent on composing clinical documents such as discharge summaries while ensuring accuracy and completeness of documentation is an important effort. Such endeavors would be particularly useful in resource-limited countries such as Kenya where there is only 1 physician per 5,000 habitants [Bank, 2014]. Fortunately, advancements in technology, as exemplified by newer approaches in recurrent neural networks, present an opportunity for the adoption of computer-generated textual summaries of patient histories which can be "as accurate as, and more efficient than, human-produced patient records for clinicians seeking to accurately identify key information about patients overall history" [Scott *et al.*, 2013]. We expect that helping physicians to minimize the paperwork can increase their time spent in meaningful interactions with patients and the ability to provide high-quality care.

Here we present a preliminary pipeline to generate discharge summaries automatically. We detailed a couple of deep learning models for abstractive summarization, trained them, tested them on the MIMIC III ICU dataset, and compare them with the state-of-the-art extractive and abstractive summarization methods.

2 Related Work

There have been many efforts to summarize EHR using Natural Language Processing (NLP) approaches. In general, NLP methods can be classified as abstractive or extractive, where extractive summaries are created by borrowing phrases from the original text, and abstractive summaries generate new text that synthesizes the original text [Pivovarov and Elhadad, 2015]. Among all summarization methods for EHR processing, extractive methods of structured data dominate the field. These are often used for summarizing tabular clinical variables from the EHR, or structured and edited-annotated textual data. Here, we study extractive and abstractive approaches for discharge summaries.

In the domain of clinical summarization, typical extractive approaches identify pieces of structured EHR data and display them in a textual format without providing additional layers of abstraction. On the other hand, abstractive summaries are more human-like and, therefore, easier to read for providers and patients because they synthesize new text from the original text and use language that is familiar to the reader [Sutskever *et al.*, 2014; Nallapati *et al.*, 2016a;

*Contact Author

Nallapati *et al.*, 2016b; Chopra *et al.*, 2016; See *et al.*, 2017; Cintas *et al.*, 2019].

Discharge summaries are generated from multiple textual clinical notes using real non-edited unstructured clinical notes. None of the current summarization methods can generate effective discharge summaries due to the following challenges [Allahyari *et al.*, 2017; Pivovarov and Elhadad, 2015; Cintas *et al.*, 2019]. First, preprocessing on raw non-annotated datasets is much more complicated than preprocessing on annotated datasets [Ganesan and Subotin, 2014; Carvalho and Curto, 2014; Curto *et al.*, 2016]. Second, most methods focus on the generation of single sentences or paragraphs from one document as oppose to multiple documents from different health providers [Liang and Tsou, 2019; Cintas *et al.*, 2019]. Third, current methods do not take into account the overall context of a patient’s visit; most of them extract data from the structured patients’ history and not from the clinical notes from the hospital stay [Allahyari *et al.*, 2017; Pivovarov and Elhadad, 2015; Liang and Tsou, 2019].

3 Methods

Figure 1 shows a high-level representation of our summarization approach, which takes as input clinical notes for a given patient and then outputs an automatically populated hospital discharge summary. Our approach has four modules. First, a module is used for preprocessing clinical notes. Second, two modules are used to extract the hospital stay context. These are the patient stratification module and the provider identification module. Third, a summarization module takes as input the preprocessed text, the type of patients, and the type of providers, and generates the final discharge summary.

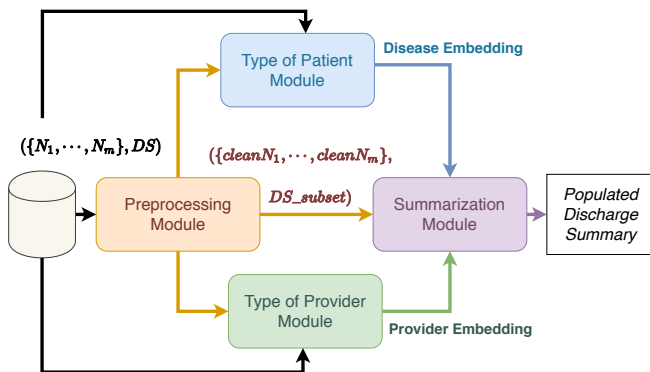


Figure 1: High-level components of the proposed approach which has four modules. One preprocessing module; two modules to extract the hospital stay context: the patient stratification module and the provider identification module; and one summarization module to generate the discharge summary.

3.1 Data Preprocessing

The preprocessing module encompasses filtering-out patients without clinical notes or discharge summaries, extracting the ‘hospital course’ free text from discharge summaries, and performing a standard text preprocessing pipeline, including lowercasing the text, replacing control characters and special

symbols, and normalizing common words. In this module, we also concatenate all the progress notes in chronological order to construct a clean set of documents (see Figure 1).

3.2 Patient Context Extraction

To extract patient context, our pipeline partitions the set of patients into a predefined number of groups (k). By grouping patients, we can get a more focused summary by feeding the model with more specialized background information versus using notes from the complete set of patients. We tested three different ways of partitioning patients: a) based on their clinical notes, b) based on their paths of care, and c) based on their diagnosed diseases. These are described below.

For the Patient similarity based on clinical notes topics context we group patients who have similar vocabularies in their clinical notes as follows. 1st, we represent each patient as a document with its clinical notes. 2nd, we compute term frequency-inverse document frequency (tf-idf) [Ramos and others, 2003] per term-document. 3rd, we reduce dimensions of tf-idf using singular value decomposition (SVD) [Golub and Reinsch, 1971]. 4th, we cluster documents based on topics using k-means clustering [Lloyd, 1982].

When we consider the context based on clinical paths of care we group patients that have similar sequences of care based on the hypothesis that patients with similar flows of care have similar discharge summaries (see Figure 2). To explain this process, we represent each patient’s path as a directed graph with the starting node being the admission unit and the ending node being a discharge from the hospital. We then measure the similarity of paths among patients by computing the edit distance between each pair of graphs [Sanfeliu and Fu, 1983]. The edit distance counts the minimum number of edit operations that should be done from one graph to another. Formally, given a set of graph edit operations, the graph edit distance between two graphs g_1 and g_2 , written as $GED(g_1, g_2)$ can be defined as:

$$GED(g_1, g_2) = \min_{(e_1, e_2, \dots, e_k) \in \mathcal{P}(g_1, g_2)} \sum_{i=1}^k c(e)$$

Where $\mathcal{P}(g_1, g_2)$ denotes the set of edit paths transforming g_1 into g_2 and $c(e) \geq 0$ is the cost of each graph edit operation e . The set of elementary graph edit operators typically includes vertex insertion, vertex deletion, edge insertion, and edge deletion. Finally, we performed hierarchical clustering [Johnson, 1967] over the graphs.

3.3 Provider Context Extraction

To extract the health provider context, our method partitions health providers by role (e.g., radiologist, general practitioner, nurse). The intuition behind being that each health provider focus on different aspects of health care and the progress notes that they write reflect this specialization. Documents written by a particular provider follow specific patterns, and identifying such patterns makes easier the abstraction of the important information that should be part of a patient discharge summary.

3.4 Context-based Abstractive Summarization

In this section, we introduce our approach to automatically generating discharge summaries. We model the hospital

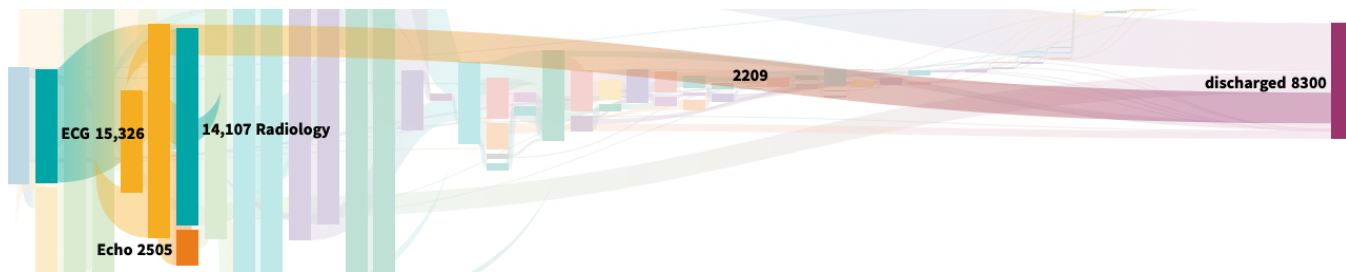


Figure 2: Example of an extracted flow of care for diabetic patients. Here, the second largest group consists of 2209 out of 8300 patients. We can observe that patient with this flow first had the sequence: (1) admission notes, (2) radiology reports, (3) electrocardiogram reports (ECG), echocardiogram reports (Echo), or radiology reports, and (4) discharged from the ICU after an electrocardiogram report (ECG).

visit **context** using neural sequence-to-sequence (seq2seq) [Sutskever *et al.*, 2014] learning with attention policies [Cho *et al.*, 2015]. This has been tested before in other domains [Luong *et al.*, 2015]. We define two implementations of the integration of the patient’s context. Model A incorporates an encoder for the type of patient where we define the patient context as all the notes coming from similar patients (see Figure 3). Model B implements a seq2seq with attention mechanism and we define the top-N topics obtained from Latent Semantic Analysis (LSA) embeddings. In the following subsections, we provide a brief review of the underlying formalisms to make this research self-contained.

Sequence to Sequence Model

At a high level, a seq2seq structure contains an encoder and a decoder to transform a sequence of text, $x = (x_1, x_2, \dots, x_n)$, from the source embedding to the desired target embedding $y = (y_1, y_2, \dots, y_m)$ [Sutskever *et al.*, 2014]. In this case, n and m represent the number of tokens (length of text) in the input and output spaces respectively. The encoder is fed the input sequence of tokens from a source document (values of x), in this study a set of clinical notes, and transforms these into a hidden state $h = (h_1, h_2, \dots, h_n)$. This hidden state h is used as input to the decoder, which then generates a summary $o = (o_1, o_2, \dots, o_m)$. The target embedding is used to define the embedding learned during the training of the summarization model. The output summary (o) is trained to match the ground truth provided during training (y) [Shi *et al.*, 2018].

Attention Mechanism

The attention mechanism [Cho *et al.*, 2015] improves the outcome of classic seq2seq models for several NLP tasks such as translation [Bahdanau *et al.*, 2014; Luong *et al.*, 2015] and summarization [Nallapati *et al.*, 2016a; See *et al.*, 2017] by selectively focusing on parts of the source sentence during the decoding phase. This results in a clear improvement of the ROUGE score compared to a seq2seq baseline model without attention mechanism [Nallapati *et al.*, 2016a; See *et al.*, 2017]. Under the attention mechanism, the decoder uses the encoded representations (h_i) for $i = 1 \dots n$ of the source description of the clinical trial but also has weights that indicate the part in which the model should focus.

In the traditional seq2seq model, we have one bidirectional recurrent neural network (RNN) as the encoder and one RNN

as the decoder with beam search to improve the model performance. To incorporate the hospital stay context, Model A has an additional RNN with attention network, see Figure 3. So instead of having one bidirectional RNN encoder, we have a stack of networks based on the type of patient and provider. This implies that after finding the type of patients, we grouped them together and then we enrich the embedding of this network with the diagnosis type. Both with attention mechanisms to improve the outcome of the model. Under the attention mechanism, the decoder not only uses the encoded representations (h_i) for $i = 1 \dots n$ of the source description of the clinical trial but also has weights that indicate the part in which the model should focus on. Here, a bidirectional gated recurrent unit (GRU) was used as the encoder and a directional GRU as the decoder [Cho *et al.*, 2014].

4 Experiments and Results

This section describes our process for training and testing the proposed approach. We trained and tested our approach using clinical notes (N_m) and discharge summaries (DN) from MIMIC-III [Johnson *et al.*, 2016], (see Figure 1). We implemented a minimal version of our approach with a preprocessing module, a topics-based patient extraction module, and two variations of the summarization module (Model A context encoder and Model B LSA embeddings). We compare our proposed models with a baseline extractive summarizer (LSTM) [Isonuma *et al.*, 2017] and an abstractive summarizer (vanilla seq2seq) [Sutskever *et al.*, 2014].

4.1 MIMIC III

The Medical Information Mart for Intensive Care - MIMIC III dataset [Johnson *et al.*, 2016] is a freely accessible critical care database widely used for healthcare benchmarking. MIMIC III consists of 40 different tables with EHR data for 46 thousand patients. It has 2 million clinical notes and 255 different diagnoses. On average, each clinical case contains 88 clinical notes, 133 tokens in the discharge summary, and 4810 tokens in clinical notes. The free text documents in mimic are unstructured, unedited, and have no annotations or levels. Preprocessing MIMIC clinical notes has many challenges that have been extensively documented in the literature [Curto *et al.*, 2016; Carvalho and Curto, 2014]. In addition to typical NLP challenges such as the text containing grammatical errors, typographical errors, and missing or misplaced

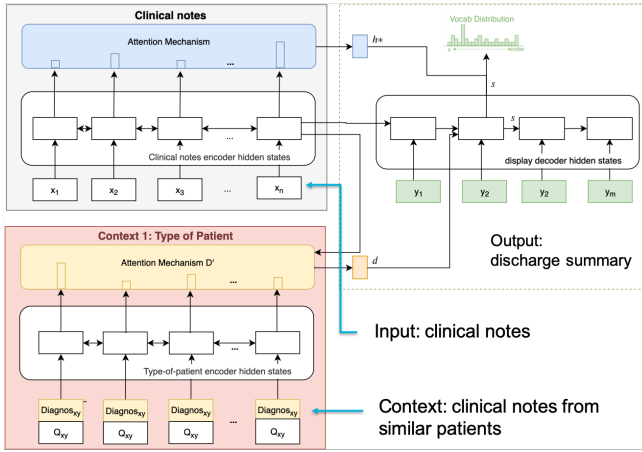


Figure 3: Model A. Summarization Module using notes of patients with similar disease as context

control characters, MIMIC texts lack function words and contain many medical terms, technical abbreviations, and numerical values associated with physiological variables readings.

4.2 Cohort Selection and Data Preprocessing

The MIMIC dataset contains EHR from 46,520 patients from where we selected the 41,127 patients that have at least one discharge summary and clinical notes. Out of these, we chose discharges and notes that correspond to the first hospital admission only and discarded cases with addendums to the discharge notes. We then selected patients with a primary diagnosis of diabetes mellitus or hypertension (ICD-9 codes 25000 - 25093 or 24900 - 24991). We implemented our approach in three modules: preprocessing module, Type-of-patient Module, and Summarization module. All the software is written in Python with the IBM PyTorch-seq2seq framework¹ for easy implementation. After preprocessing, the average number of tokens in the concatenated clinical notes is 4,810 and 133 tokens in discharge summaries.

4.3 Performance evaluation

After preprocessing, the discharge summaries dataset is composed of 1,394 pairs, which was split into a training subset containing 976 pairs (70%), a validation set of 140 pairs (10%) and a testing set of 278 pairs (20% of the total dataset) that were selected with a random permutation cross-validation iterator. To evaluate the performance of our models, we summarize the clinical reports using a LSTM-based extractive summarizer as one baseline [Isonuma *et al.*, 2017; Hochreiter and Schmidhuber, 1997] and vanilla seq2seq abstractive summarizer as a second baseline [Sutskever *et al.*, 2014]. The baseline Long short-term memory (LSTM) approach uses an LSTM model for sentence extraction and classification for single-document summarization [Isonuma *et al.*, 2017; Hochreiter and Schmidhuber, 1997].

To evaluate the generated summaries, we used the ROUGE metric for which given the ground truth discharge summary and the generated summary, we find the percentage of n-gram from the ground truth summary appear in the automati-

¹<https://github.com/IBM/pytorch-seq2seq>

	ROUGE-2	ROUGE-1	ROUGE-L
LSTM	0.30	0.37	0.27
Seq2seq	0.76	0.83	0.83
A.Encoder	tr: 0.94 te: 0.05	tr: 0.95 te: 0.23	tr: 0.95 te: 0.19
B.Embedding	tr: 0.87 te: 0.05	tr: 0.90 te: 0.23	tr: 0.90 te: 0.20

Table 1: ROUGE-2, ROUGE-1, ROUGE-L F-score over the summarization models. The first row shows the results of the extractive summarization baseline LSTM. The second row shows the results of the abstractive summarization baseline Vanilla seq2seq. The third row shows the partial results of the proposed Model A – adding a context encoder. The fourth row shows the partial results of the proposed Model B – adding the top-N topics from LSA embeddings.

cally generated summary [Lin, 2004]. We compared bigrams (ROUGE-2), unigrams (ROUGE-1), and the Longest Common Subsequence (ROUGE-L).

4.4 Preliminary Results

Table 1 shows the preliminary results when testing on the 1,394 progress notes that have a source-target ratio of 50 to 4 words. Model A, which uses clinical notes of similar patients as context encoder (see Figure 3), achieves a ROUGE-2 of 0.94, ROUGE-1 of 0.95, and ROUGE-L of 0.95 for training, but very low scores for testing which shows that our model is over-fitting. Model B, which uses the top topics from LSA embeddings as context, achieves a ROUGE-1 of 0.87, ROUGE-1 of 0.90, and ROUGE-L of 0.90 for training.

To extract patient context, we stratify diabetes patients based by representing them as graphs of events from their admission to discharge. We identified the different flows of care among patients using Sankey plots for visualization (as illustrated in Figure 2) and cluster patients based on their flows of care. We compare the two methods for clustering patients for the type-of-patient module using Silhouette Coefficient. The clustering of patients based on paths of care has a Silhouette Coefficient of 0.203 and clustering based on topics 0.312.

5 Conclusions and Future Work

By leveraging patients’ hospital stay context extracted from progress notes, we propose an attention-based approach that can accurately generate discharge summaries from the MIMIC dataset. We credit this achievement in large part to the type of patient module which characterizes patients based on their flow of care.

Our preliminary results suggest that our approach has the potential to be adopted as a tool to assist health professionals in the generation of discharge summaries and that it captured the context of the patient’s hospital stay. Although these results are promising, we recognize that further experiments with larger and heterogenous datasets need to be done, and additional validation by human experts is required. We also hypothesize that a hybrid approach that includes health professionals and a neural network would be ideal for this sensitive task. We are working towards the implementation of the provider context, which we believe will increase the performance and explainability of the model.

References

- [Allahyari *et al.*, 2017] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D Trippe, Juan B Gutierrez, and Krys Kochut. Text Summarization Techniques: A Brief Survey. *arXiv:1707.02268 [cs]*, 2017.
- [Bahdanau *et al.*, 2014] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- [Bank, 2014] The World Bank. Physicians per 1,000 people. <https://data.worldbank.org/indicator/SH.MED.PHYS.ZS?locations=KE>, 2014.
- [Carvalho and Curto, 2014] Joao Paulo Carvalho and Sérgio Curto. Fuzzy preprocessing of medical text annotations of intensive care units patients. In *2014 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*, pages 1–7. IEEE, 2014.
- [Chaboyer *et al.*, 2008] Wendy Chaboyer, Lukman Thalib, Michelle Foster, Carol Ball, and Brent Richards. Predictors of adverse events in patients after discharge from the intensive care unit. *American Journal of Critical Care*, 17(3):255–263, 2008.
- [Cho *et al.*, 2014] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
- [Cho *et al.*, 2015] Kyunghyun Cho, Aaron Courville, and Yoshua Bengio. Describing multimedia content using attention-based encoder-decoder networks. *IEEE Transactions on Multimedia*, 17(11):1875–1886, 2015.
- [Chopra *et al.*, 2016] Sumit Chopra, Michael Auli, and Alexander M. Rush. Abstractive Sentence Summarization with Attentive Recurrent Neural Networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016.
- [Cintas *et al.*, 2019] Celia Cintas, William Ogallo, Aisha Walcott, Sekou L Remy, Victor Akinwande, and Samuel Osebe. Towards neural abstractive clinical trial text summarization with sequence to sequence models. In *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–3. IEEE, 2019.
- [Curto *et al.*, 2016] Sérgio Curto, Joao P Carvalho, Cátia Salgado, Susana M Vieira, and João MC Sousa. Predicting icu readmissions based on bedside medical text notes. In *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 2144–a. IEEE, 2016.
- [Ganesan and Subotin, 2014] Kavita Ganesan and Michael Subotin. A general supervised approach to segmentation of clinical texts. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 33–40. IEEE, 2014.
- [Golub and Reinsch, 1971] Gene H Golub and Christian Reinsch. Singular value decomposition and least squares solutions. In *Linear Algebra*, pages 134–151. Springer, 1971.
- [Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [Isonuma *et al.*, 2017] Masaru Isonuma, Toru Fujino, Junichiro Mori, Yutaka Matsuo, and Ichiro Sakata. Extractive summarization using multi-task learning with document classification. In *Proceedings of the 2017 Conference on empirical methods in natural language processing*, pages 2101–2110, 2017.
- [Johnson *et al.*, 2016] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3:160035, 2016.
- [Johnson, 1967] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.
- [Kripalani *et al.*, 2007a] Sunil Kripalani, Amy T Jackson, Jeffrey L Schnipper, and Eric A Coleman. Promoting effective transitions of care at hospital discharge: a review of key issues for hospitalists. *Journal of hospital medicine: an official publication of the Society of Hospital Medicine*, 2(5):314–323, 2007.
- [Kripalani *et al.*, 2007b] Sunil Kripalani, Frank LeFevre, Christopher O Phillips, Mark V Williams, Preetha Basaviah, and David W Baker. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. *Jama*, 297(8):831–841, 2007.
- [Liang and Tsou, 2019] Jennifer Liang and Ching-Huei Tsou. A novel system for extractive clinical note summarization using ehr data. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 46–54, 2019.
- [Lin, 2004] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81, 2004.
- [Lloyd, 1982] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.
- [Luong *et al.*, 2015] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.
- [Nallapati *et al.*, 2016a] Ramesh Nallapati, Bing Xiang, and Bowen Zhou. Sequence-to-Sequence RNNs for Text Summarization. *CoRR*, abs/1602.06023, 2016.
- [Nallapati *et al.*, 2016b] Ramesh Nallapati, Bowen Zhou, Caglar Gulcehre, Bing Xiang, et al. Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*, 2016.

- [Pivovarov and Elhadad, 2015] Rimma Pivovarov and Noémie Elhadad. Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association*, 22(5):938–947, 2015.
- [Ramos and others, 2003] Juan Ramos et al. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, volume 242, pages 133–142. Piscataway, NJ, 2003.
- [Sanfeliu and Fu, 1983] Alberto Sanfeliu and King-Sun Fu. A distance measure between attributed relational graphs for pattern recognition. *IEEE transactions on systems, man, and cybernetics*, 1(3):353–362, 1983.
- [Scott et al., 2013] Donia Scott, Catalina Hallett, and Rachel Fettiplace. Data-to-text summarisation of patient records: Using computer-generated summaries to access patient histories. *Patient education and counseling*, 92(2):153–159, 2013.
- [See et al., 2017] Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. *CoRR*, abs/1704.04368, 2017.
- [Shi et al., 2018] Tian Shi, Yaser Keneshloo, Naren Ramakrishnan, and Chandan K Reddy. Neural abstractive text summarization with sequence-to-sequence models. *arXiv preprint arXiv:1812.02303*, 2018.
- [Sinsky et al., 2017] Christine Sinsky, Michael Tutty, and Lacey Colligan. Allocation of physician time in ambulatory practice. *Annals of internal medicine*, 166(9):683–684, 2017.
- [Sutskever et al., 2014] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.
- [Woloshynowych et al., 2003] M Woloshynowych, G Neale, and C Vincent. Case record review of adverse events: a new approach. *BMJ Quality & Safety*, 12(6):411–415, 2003.