Analyzing Narratives of Patient Experiences: A BERT Topic Modeling Approach

Mátyás Osváth[1], Zijian Győző Yang[2,3], Karolina Kósa[1]

[1]Department of Behavioral Sciences, Faculty of Medicine,
University of Debrecen
Nagyerdei krt. 98, H-4032 Debrecen, Hungary
osvath.matyas@med.unideb.hu; kosa.karolina@med.unideb.hu

[2]Hungarian Research Centre for Linguistics
Benczúr u. 33, H-1068 Budapest, Hungary
yang.zijian.gyozo@nytud.hu

[3]Sightspot Network Ltd.
Vágóhíd u. 2, H-4034 Debrecen, Hungary
yang.zijian.gyozo@nytud.hu

Abstract: Due to healthcare systems increased focus on healthcare quality and patient-centered care, the patients' perspective of delivered healthcare, has become an important part of healthcare service evaluations. Patient experiences can be used to improve the quality of care, as they reveal important information about health care encounters. An increasing number of organizations systematically collect and analyze patient experience data. The aim of our study was to identify major topics in narratives of patients' healthcare related experiences and analyze the reactions of readers of patient experiences. 1663 blogs and 298806 textual comments were extracted on non-solicited patient experiences from a Hungarian online forum during a 10-year period. Topic modeling with state-of-the-art BERT embeddings were used to analyze the data and extract meaningful patterns and concepts. Sentiment analysis was utilized to categorize the emotional valence of the narrative writings. The huBERT and HIL-SBERT models identified 326 and 200 topics in terms of patient experiences and 508 and 728 topics regarding the reactions to these experiences without human supervision. Conceptually similar topics were integrated into major categories with manual analysis. 94.4% of the experiences and 77.5% of comments were classified as negative, reflecting the same annual tendency over the decade. Our study uses a data-driven approach for extracting patterns of healthcare related patient opinions, in Hungary. Topic modeling, based on BERT embeddings, could provide useful information on patient perceptions and perspectives, that could improve healthcare quality and safety.

Keywords: NLP; topic modeling; sentiment analysis; patient experience; health care quality

1 Introduction

One of the indicators of health system performance is healthcare quality, which is high priority for the health policy sector and healthcare organizations aiming at patient-centered and quality care. Several theoretical frameworks – consisting of various indicators – have been developed for measuring healthcare quality. In 1966, Avedis Donabedian published a lasting framework and proposes to characterize the quality of healthcare with indicators that can be classified into three dimensions, namely structure, process and outcome [1]. The Organization for Economic Co-operation and Development (OECD) issued a framework as a nested-matrix aimed to measure the healthcare system performance, further developing Donabedian's quality of care [2]. One of the matrices essential dimension is patient experience, which includes every interaction between patient and the system healthcare, influencing the patients' opinion about healthcare. Amongst others, patient experience, comprises the communication between patient and healthcare personal, access to care and waiting time. It is important to distinguish "patient experience" from a different indicator called "patient satisfaction", which refers to patients' evaluation of their care compared to what they previously expected [3].

Along with the proliferation of the Internet, sharing experiences of and opinions on health services have "skyrocketed" in the form of reviews, blogs and comments. Patient online reports (POR) and patient report websites (PRW) allow users to review their or others health care encounters freely and anonymously in the form of quantitative ratings and/or textual feedbacks [5]. Analysis of non-structured texts and reviews can contribute to the improvement of healthcare quality and advance in person-centered measures [6]. Online reviews and comments could affect the attitudes and perceptions of other users and influence the decision making of patients [7] when selecting physician and healthcare services. According to a Hungarian study, at least 60% of the respondents chosen a physician based on online information and 17% of them regularly seek medical information online [8], which number is increasing.

Organizations increasingly collect and analyze online data anonymously with APIs and web scraping techniques. In the USA, several organizations operate web applications (RateMD, Healthgrades, Yelp) to collect and analyze patients' quantitative ratings and qualitative textual feedbacks describing the received care in a certain hospital and department. The British National Health Choices (NHS) maintains an online portal for writing narrative feedbacks [9]. In Hungary, studies mostly used such survey methods to assess patients' experiences [4]. Studies suggests an association between positive patient experiences with less frequent use of secondary and tertiary care (e.g., hospital admission and readmission, consultations) [10]. Improved adherence, lower mortality, lower readmission rates and shorter inpatient stays characterizes hospitals with higher patient satisfaction scores [6].

Although the application of quantitative methods is dominant in measuring patient satisfaction and experiences, these methods are necessarily limited due to information loss and less intelligible and practical to healthcare personal. Gallan et al. (2016) concluded discrepancy between ratings and comments, i.e., highlighting the possible contrast between survey scores and textual feedbacks [11]. Text analysis allow a more detailed and fine-grained processing but requires more complicated techniques. Manual content analysis and annotation has been the standard way to analyze narratives in any domain [9] [12] [13], but practically impossible to extract information from large quantities of text. In the past decades, progression in Natural Language Processing (NLP) made the analysis of large volumes of text not only possible but scalable and proved to be an alternative to be used for information extraction and identifying valuable patterns from non-structured texts.

1.1 Related Work

Applications of Machine Learning (ML) and Natural Language Processing (NLP) to data related patient experiences face several challenges, for instance the volume, type of corpus, classification models used and overfitting [14]. Greaves et al. [15] utilized sentiment analysis and machine learning to classify patients' comments, and ML algorithms to predict different categories, such as hospital cleanliness, whether patients were treated with respect, and hospital recommendation. Doing-Harris et al. [16] extracted latent topics in patient comments about health care services using a machine learning based approach. Among others, their most frequent categories were related to appointment access, empathy, and explanation. Li et al. [5] developed a taxonomy for physician reviews including 3 domains related to physicians, patients and systems as well as 9 subcategories. A mixed-method approach was used including literature review, human annotation, and topic modeling.

In 2021, a systematic review summarized studies that applied ML and NLP models to patient experience to identify patterns from non-structured texts [12]. Studies were categorized based on supervised, unsupervised, semi-supervised, rule-based or dictionary-based approaches. As the study highlighted, a common method utilized is sentiment analysis, which is used to detect emotional valence expressed by the writer in the text, classifying texts into negative, positive or neutral category. All studies with unsupervised approaches (n=6) used topic modeling, i.e., to discover latent semantic structure and automatically identify topics in a collection of documents and texts. Different algorithms can be used to recognize patterns and give new insights within data. However, most studies relied on Latent Dirichlet Allocation (LDA) algorithm for topic modeling, which – according to Angelov [17] – has several disadvantages including the need to define the number of topics a priori and disregarding word semantics. We find one study, where contextual information from clinical texts of patients – such as disease severity, time and onset

– were extracted by using state-of-the-art BERT embeddings. Their experimental result achieved higher accuracy rate than with Word2vec model [18].

According to our knowledge, this is the first study in Hungary that focuses on analyzing narrative patients' experiences as well as reactions to these experiences by other readers.

1.2 Objective

We hypothesized that non-structured blogs and users' responses to them in the form of comments provide meaningful information on patient experiences, which can be extracted by NLP techniques, getting insight from patients' and users' texts. The aim of the study was to identify patterns and topics in non-solicited patient experiences about Hungarian healthcare system posted to an online forum. Furthermore, readers' comments on the spontaneous narrative experiences were examined to understand the interactions in the comment section. To encode the text of the comments, two state-of-the-art contextualized BERT models were applied, and unsupervised machine learning algorithm were utilized to cluster the results. Additionally, sentiment analysis (with a fine-tuned BERT model) was used to categorize the valence (positive, negative, and neutral) of non-structured texts.

2 Methods

2.1 Data Source

Data were extracted from an online forum (https://praxis.blog.hu/) dedicated to sharing patient experiences in Hungarian health care. The non-institutional website was established in 2009, and the daily number of visitors ranges from 5000 to 7000. Spontaneous descriptions of health care encounters emailed to the staff by patients, or their relatives are posted without editing anonymously in the form of blogs. Contents on the website are organized by week and year. Anyone can comment under each post and answer to a comment. Blogs and comments do not have word limits. Blogs are uploaded voluntarily and anonymously so that the identity of the writer or persons depicted in the blog cannot be established. Approval and informed consent were not necessary as the website from which blogs were downloaded is accessible to anyone and can be freely used for non-commercial purposes by referencing the source according to the Creative Commons License of Hungary (CC BY-NC-ND 2.5 HU) [38].

Our research focuses on the blogs and comments, with a total of 1663 blogs and 298.806 comments on patient experiences posted between 2009 and 2020. Data were extracted by a Web crawler, and we deleted any information on personal identity or sensitive data. The final cleaned dataset comprises 1660 blogs and 267631 comments.

2.2 Data Analysis

Word embeddings were used to convert comments into numerical data. These are vector representation of words in a text, capturing contextual hierarchy. From that, similarity metrics can be calculated using distance measures (e.g., cosine distance). Deep learning models, such as Word2Vec are often used for extracting features and capturing context from corpus, but problems may arise in a highly specific context. Bidirectional Encoder Representations from Transformer (BERT) [19] resolves this issue by learning the contextual features between words, embedding into a continuous vector space, reading input text sequentially from right-to-left and left-to-right. BERT models can be pre-trained on large corpora of text and fine-tuned to be used in different contexts.

Our study builds on the prior work of Angelov [17] which applies state-of-the-art Transformer based word embeddings for topic modeling. We used two contextualized BERT embeddings to encode the text of the comments. For the first BERT model, namely huBERT, preprocessing of the input text were not necessary, except for tokenization, as the model was trained similarly on a 9-billion-token Hungarian Webcorpus 2.0 [20], outperforming other BERT models in various NLP tasks [21]. The architecture of huBERT is the same as BERT base model with 12 Transformer layers with 12 heads and 768 hidden dimensions [22]. Additionally, a fine-tuned HIL-SBERT model [23] was also applied for feature extraction which is based on Sentence-BERT [24], a modification of BERT model aiming to achieve high-quality sentence embeddings.

To improve the results of unsupervised clustering algorithm used for topic identification, dimension of the data was reduced using the Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) algorithm [25]. Next, the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm [26] was applied to the encoded low dimensional vector representation to capture the structure of the data automatically with the aim of grouping words to form topics. HDBSCAN is a density-based hierarchical clustering algorithm that does not require pre-defined number of clusters providing optimal solution for the reduction of problems with vector sparsity and computation costs. Models were fine-tuned to have 15 words per topic. Clustered results were lemmatized, and stop words and punctuation marks were removed for easier interpretation. Topics with null weights and duplicated topics were removed from the output. For interpreting the clustered results, c-TF-IDF (class-based TF-IDF) [27] were used for automatic topic identification. The final output of the pipelines were clusters, consisting of 15 words per cluster, and human interpretation was used to assign labels for the topics and merged some of the topics to form more interpretable categories.

Sentiment analysis was utilized to identify the emotional content of the comments. To fine-tune the pre-trained huBERT model, the Hungarian Twitter dataset [28] – created by PreCognox [29] – was used. Two subcorpora were created as defined in the international benchmark [30] and applied to predict the class label of comments:

2-class: Class label 1 and 2 were converted to negative, class 4 and 5 to positive, ignoring 3 for ambiguities. Training corpus: 2468 segments. Test corpus: 269 segments.

3-class: Class label 1 and 2 were converted to negative, class 4 and 5 to positive and class 3 to neutral. Training corpus 3600 segments. Test corpus: 400 segments.

5-class: Classes were labelled as in the original Likert scale with 1 (very negative) to 5 (very positive).

Fine-tuning was performed with the following hyperparameters: batch size 8 / GPU (4 GeForce GTX 1080Ti); learning rate: 2e-5; sequence length of 128 and 15 epochs (models with best performance were used). The model achieved an accuracy of 85.92% for the binary, 72.18% for the 3-class corpus and 68.50% for the five-class variant.

3 Results

3.1 Topics Modeling using Patient Experiences

Analysis revealed that the resulting clusters with fine-tuned BERT models form semantically adequate topics. The encoded patient experiences analyzed by the huBERT and HIL-SBERT models resulted in 326 and 200 topics, respectively. Both models only use part of the corpus, 20503 sentences by the huBERT, and 20502 by HIL-SBERT model, clustering the rest as noise (huBERT: 33744, HIL-SBERT: 34745). Dimensionality reduction was applied to visualize the clustered results and show the representation of comments in continuous vector space.

No quantitative metrics were found to evaluate the unsupervised models; therefore, results were interpreted manually. The outcome of huBERT model had easier interpretation, i.e., resulted in more interpretable and semantically meaningful categories, while HIL-SBERT model had more noise in the results. Therefore, during the rest of the analysis we used to topics identified by the huBERT model. Irrelevant topics were discarded from the 326 topics. Human interpretations were applied to merge topics into higher level categories. To reduce bias, authors discussed topics iteratively. Sentences in patient experiences are highly dependent to the specific situation in healthcare and to additional topics in the patient experience, as a narrative could consist of many topics. Considering their large number, only the five most notable categories will be described below, along with 18 topics marked important by the model (see Tables 1 and 2).

Diagnosis, Symptoms and Illnesses

This category comprises patients' or their relatives' perceived symptoms, illnesses along with their diagnosis. Symptoms were related to acute and chronic diseases ranging from influenza to cancer. Multiple topics were included in this category, as the model recognized multiple classification of diseases. This category captured dental, spine and ophthalmic problems, sexually transmitted diseases and plastic surgery. Expressions related to pain may indicate the difficulty of their condition and coping with pain. One patient wrote that: "…, pain was unbearable, as they did not anesthetize me…", another patient mentioned: "The diagnosis was a suspicion of purulent meningitis, probably correctly…"

Birth and Gynecology

Experiences related to gynecology, childbirths and birth were well differentiated by the BERT model. Recurring occurrences of topics has grounded the grouping of topics into one category. Category included the following topics: gynecological examinations, childbirth, pregnancy, abortion, infant, new-born, vaccination and protection. The model also identified topic related to postpartum depression. Expressions in this category included sentences like "Paradoxically, postpartum period doesn't bring happiness into mothers' lives." and "The child-birth itself went smoothly, the doctor was fair."

Family and Children

This category referred to the topics, which were likely to appear in the lives of families including hospitalization of the child, caring for the child, family and parenting roles, vaccination, and the condition of other relatives. It should be highlighted that several blog posts also reported events at the school and with the school's physician. The reason for separating this category from the former is that the former focuses on the mother, while words related to other relatives and children appear more frequently in the latter topics. One relative expressed: "A child must be treated with patience and attention, especially if he is ill."

Structure

Patients discussed multiple topics which can be classified into structure category forming the basis of the healthcare system and is consonant with Donabedian's dimension of structure. Topics identified by the model include problems with professionality, lack of human resources, relationship between healthcare and politics, and the malfunctioning of healthcare system. Multiple references were made regarding the inadequate hygiene circumstances in the hospitals and departments. One patient articulated as follows: “I was thinking of sharing this with you because now there is chaos, bankruptcy, lack of physicians everywhere.”

Outcomes

Both favorable and unfavorable expression related to outcomes were included in this category. Sentences in unfavorable outcomes comprise dying, death and mourning and the uncertainty of recovery. The model identified several inauspicious outcomes of healthcare. One relative wrote the following: "The family believes there was an omission in the hospital and the overdose of medication led to the death of our loved one." Favorable outcomes included expression related to topics like healing and gratitude, or references to quality care provided by the healthcare personnel. One patient expressed: “A thousand thanks to those who are human despite all the difficulties.”

Table 1
Meaningful categories of patient experiences

#	Category	Keywords
1	Diagnosis, symptoms and diseases	spine, waist, toothache, herpes, pain, disc herniation, infusion, complaint, sclerosis, numbness, double vision, muscle pain
2	Structure	lack of nurses, disorganization, lack of doctors, remorse, care system, unfair, professional, heedlessness, high degree,
3	Outcome	work, class, grateful, love, nurse, working, bond, heart, mourning, man, mourning, death, dead, process
4	Birth and gynecology	mom, gynecologist, crib, childbirth, childbirth, abortion, slut, vaccination, protection
5	Family and children	toddler, bed, intelligence, need, development, child, father, role, mother, tongue, parent, carrier

Table 2
Keywords and their classification

#	Topic	Keywords
1	Medical Oath	Oath, Hippocratic, if, medical, worthy, insight, part, recognition, fulfil
2	Medicine and Drugs	Frontin, hormone, escitil, ingredient, acetyl, sore throat, ibuprof, tablet, paracetamol, agent, tablet
3	Vaccination	Vaccination, vaccination, against, influenza, protection, vaccination, pharmacy, administration
4	Blood	Blood, blood donation, blood donation, call, potential, blood pressure
5	Alternative medicine	Magneter, long, effect, initial, alternative, deterioration, cure, various, unexpected, improvement
6	God and science	God, eso, science, extreme, medicine, esoteric, pedestal,
7	Plastic intervention	Plastic, horrible, surgery, chest, detail, week, order, tube, cm, pair, surgeon, breast augmentation

8	Other healthcare services	Soothing, dispatcher, ambulance, smile, patience, sick, car, soothe, nurse, professional
9	Holidays and religion	Wreath, advent, candle, stove, Christian, mood, mythology, decorate, pine tree
10	Healthy lifestyle	Diet, food, dietary, varied, seven, nutritional, magnesium, stress, workplace, body, magnesium deficiency, b6, stressful
11	Workout	Workout, movement, amplifitness, proper, training plan, fitness room, muscle group, body awakening, device
12	Side effect(s)	Side effects, truxal, generalized, rulid, psychotic, paranoid, anxious
13	Smoking	Smoking, act, experience, anyone, doctor, many, get used to, addiction
14	Communication about health condition	Doctor, condition, communicate, ct, brother, how many, day
15	Referral	Date, examination, request, sheet, referral, give, know, doctor
16	Sharing experience	Story, love, share, experience, tell, blog, hospital, write, describe
17	Money	Tb, pay, ft, tax, disability benefit, aid, salary, support, money
18	Waiting times and appointment	Three, seven, days, after, hour, date, get, year, surgery, test, also, takes, waiting, line, after, half, year, waiting

Additional topics are shown in Table 2. Patients and their relatives discussed many topics in relation to health and health care such as blood, vaccination, medicine and their side effects, alternative cures, religion and science, smoking, lifestyle and workout. Further research is needed to answer the various questions arisen because of the uniqueness of these topics such as the roles of alternative cures, lifestyle, and the connection regarding the role of religion in science.

3.2 Topics Related to Users' Reaction (comments)

In contrast to the blog post topics based on the experiences of patients or relatives, users' reaction to experiences of patients mainly relies on the context. Linguistic characteristics reflected in comments contains meaningful information for determining the topic of feedback and the overall reactions to blog posts. Similarly, the same topic modeling pipeline were applied for the comments with c-TF-IDF for automatic topic identification and human interpretation for further processing.

The topics of comments resulted in four main categories, including symptoms, ethical issues, addressing feelings and consolation (Table 3). Symptoms, diseases and diagnosis category were mainly similar to the aforementioned category with the same name, referring to comments acknowledging the significance of specific symptoms of bloggers and commenters. The ethical issues category consists of several subjects of ethical relevance including police, accusation and lawyer reveals a growing interest in the consequences of inappropriate healthcare, reflecting the available actions after an unfavorable outcome. The category of addressing feelings captures commenters feelings and emotions to other patients' experiences in healthcare. As different scenarios may trigger different feelings in the commenters, they may reply with anger to the causative factor or person and feel empathy toward the patient receiving inappropriate service. Expressions in consolation category mainly containing understanding and get-well messages signaling consolation and condolences. Users may show empathy and sympathy towards a patient whose encounter could end in unsatisfactory outcome.

Table 3

Meaningful categories of comments to patient experiences

#	Category	Frequent words
1	Symptoms	vision, mental disorders, chest pain, insulin, pain, virus, bacteria, cancer, bone, knee injury
2	Ethical issues	imprisonment, police, accusation, lawyer, theft, rules, protocol, patient representative, informed consent, custodial, physical, years, public interest, insult, danger, crime, violence, negligence, report, call, legal, summoned, witness evidence, criticize on personal grounds, house rule, state, justice, required, law
3	Addressing feelings	love, grief, anger, sadness, depression, rough, argument, outrageous, gas, scandal, unpleasant, fuhh, bloodthirsty, annoying, offended, back, outrageous attitude, fucking good, terrible, bad, disgusting, rude
5	Consolation	Sympathy, get well, condolences, apologies, my condolences, accept, feel, together, done happened, dear, get well, calm, rest in peace, settle down, meditation, error

3.3 Sentiment Analysis

Patient Experiences

Categorization of patient experiences by the fine-tuned huBERT model for the 10-year period revealed that the valence of sentiments was dominantly negative (94.4%) in the five-class variant. Similar patterns were showed in the other two variants. In the five-class model, the score 3 (neutral score) were used as cut-off point for deciding the polarity of comment (positive, negative). One patient experience could contain multiple emotional expressions, whereas different aspects and processes of care may lead to different patient perceptions, and it is possible that the same hospital care may result in a positive experience for one patient and a negative experience for another. The same tendency showed when segmented the analysis of the experiences by each year in the 10-year period, that is, most of the patient experiences were negative (Fig. 3).

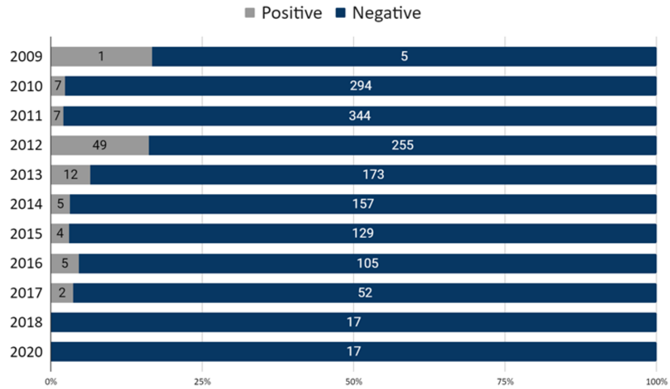

Figure 3
Sentiment analysis results during the 10-year period for patient experiences

Trend analysis were used to capture the changes in the emotional patterns in patients' experiences during the 10-year period. The sentiment score in the classified sentences were averaged and visualized sorted by year. As shown in Fig. 4, the majority of blogs reflect negative sentiment. Similar pattern can be seen in Fig. 5, which shows the distribution of the summarized sentiments of the classified sentences plotted on a five-point Likert scale.

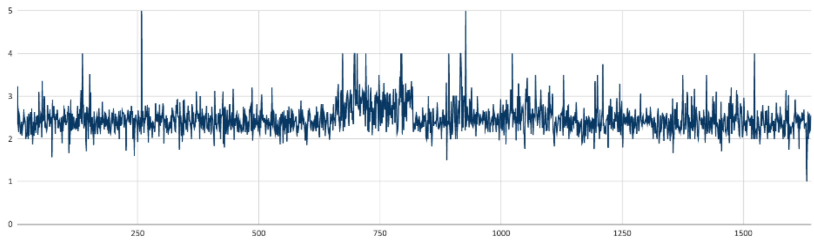

Figure 4
Distribution of blog posts by an average sentiment score sorted by years

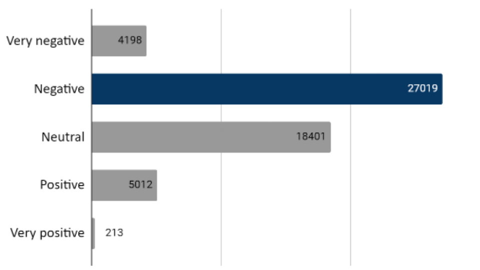

Figure 5
Overall distribution of the (blog posts') classified sentences

Comments

Similarly, the results of topic modeling showed the same tendency with the fine-tuned huBERT model for the 10-year period in the categorization of users' reactions to experiences, revealing dominantly negative (77.52%) sentiment in the binary (2-class) variant. Analysis by each year in the 10-year period showed the similar pattern, that is, the majority of sentiments in the comments were negative (Fig. 6). Commenters expressed mixed emotions (e.g., love, anger). Users may express their feelings and emotion towards the putative factors or persons that could have caused negative patient experience, not necessarily towards the specific patient, which may explain our results.

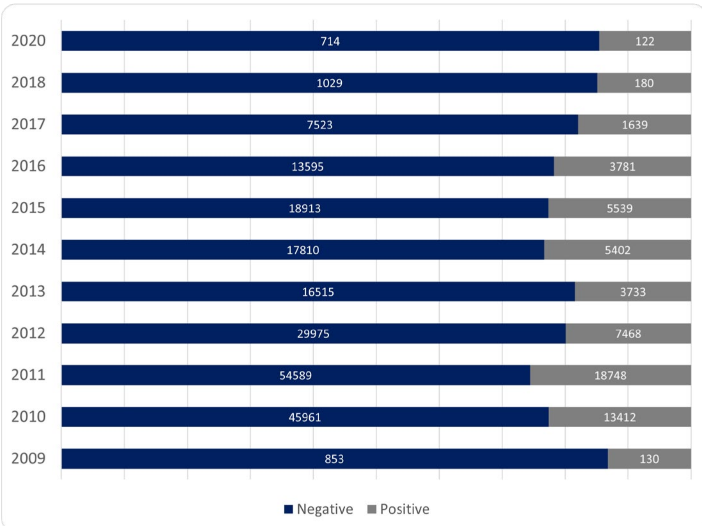

Figure 6
Sentiment analysis results segmented by year for the comments

4 Discussion

Our exploratory results present the first NLP-assisted analysis of narratives of patient experiences and reaction of various persons to narratives, emphasizing and reflecting mostly negative experiences occurring in Hungarian health care. Understanding the attitudes and opinions of patients is crucial as experiences reflects healthcare quality and may affect others' opinions and decision making related to the use of health care services. Furthermore, sharing opinions on the internet are growing and patients' views about health care services are becoming like other paying services.

Categories and topics identified in the narratives shows the frequently occurring problems healthcare considered important by patients. Our results revealed that five categories emerged from the analysis, namely symptoms and illnesses, structure, outcome of care, family and children, and birth and gynecology. Two categories are related to the classical theoretical framework describing quality of healthcare by Avedis Donabedian [1]. Other research has shown that satisfaction is correlated with system and management issues and coordination of care, for instance waiting time, access to care, hygiene and environment [9], which all contribute to negative experience if the service is unsatisfactory. The category of outcome of care were also relevant in our analysis. It is important to mention the subjective perception of care may differ between patients, as one patient could experience the treatment as suboptimal, while the other may rate the same treatment satisfactory. Additionally, other factors – such as previous illnesses, lifestyle factors, adherence – could have an impact on the outcome of care and were, "unfortunately", unknown in the blog posts.

The category of symptoms, diseases and diagnosis were prominent in both blog posts and comments, referring to the acknowledgement and significance of specific symptoms and diseases. Identifying and classifying various symptoms and diseases in narratives based on neural networks could contribute to the improvement of Clinical Named Entity Recognition systems. Moreover, a recent study proposed a framework to understand the contextual information of the parent entity (e.g., headache) to predict other symptoms' characteristics like time and severity, achieving superior performance against other state-of-the-art models [18].

Comments are mainly related to the specific context of health care encounter to which they are related. Our results show that – besides symptoms and disease category – three other categories emerged from the users' reaction to patient experiences. First, addressing feelings category contains mixed emotions. According to Table 3, there were many words related love, grief, anger and sadness. The accumulation of emotional words could help analyze the provoking factors behind negative experiences and develop a more personalized system focusing on negative patient feedbacks. Second, the category of ethical aspects and the violation of patient's right are one of the most essential factors to decrease satisfaction. The category consists of comments such as "The hospital must be sued…" or "You have to go back with a lawyer and ask for every documentation possible!".

Several topics identified by topic modeling help understand the various aspects and nuances of patient experiences and comments. Waiting times for appointment and care, delayed operations and difficulty of booking an appointment all contributes to negative experiences. According to Fernandes et al. [31] waiting times for in healthcare services differs in Hungary, the majority of respondents were seen within a week by family doctor, public specialist, or a private specialist; and only one third of patients waited more than one month. Based on our results, medicine and side-effects were important topic to patients, as many patients could have side effects after treatment [32]. Moreover, monitoring patients' feedbacks about potential side-effects, and what to do in case of specific problems develop could increase patients' satisfaction. Establishing a relation between drugs and side-effect could be included in future research.

Our results revealed that the majority of experiences (94.4%) and comments were negative (77.5%). The same tendency shows when texts were segmented by year (see Fig. 3 and 6), contrary to other studies findings [9] [13] [33]. By taking advantage of previously created Hungarian Twitter dataset, we have been able to fine-tune the BERT model and use more learning set to achieve higher level of accuracy. Commenters expressed similar feelings as in the patients' writings, reflecting the negative experiences in health care. Nonetheless, the same user can react both positively and negatively to a specific patient experience, given that it could comprise multiple healthcare encounters, it makes more difficult for the sentiment algorithm to provide accurate results.

To our knowledge, one previous Hungarian study assessed patient experiences on a national representative sample (n=1000), which included several aspects of care highlighted by the OECD framework including patient-doctor communication, shared decision-making, time spent with patients and giving opportunity to ask questions. According to their research, all categories were dominantly positive (>80%), but lower when compared to other OECD countries [4] [31].

Our analysis showed a much higher proportion of problems with healthcare services (e.g., structure and waiting times). Although our source platform is skewed towards unsatisfied and frustrated patient experiences and they are do not accurately present the quality of care in Hungary, it proved to be a useful information source as it emphasizes the categories and topics which patients finds useful.

A psychological phenomenon, namely the negative bias could further explain why patients with unsatisfied experience are more likely to write about their frustration. Negative experiences are perceived to be more complex than positive ones and therefore result in more complex cognitive representations compared to positive ones. Since negative stimuli carry greater informational and adaptive value than positive stimuli, patients spend more time dealing with negative rather than with positive stimuli [34]. The majority of negative sentiments may refer to that those patient post on this online forum that could not or do not want to submit a complaint to patients' representatives in the legal way that exists in healthcare. Nevertheless, negative experiences provide greater informational value compared to positively valenced or neutral experiences because inferences can be drawn for the improvement of health care quality.

As to the limitations, it is important to mention that our source of data is not representative of the Hungarian healthcare, therefore generalizability is limited for drawing conclusions. Unsolicited online reviews have a higher probability of selection bias as unsatisfied patients are more likely to complain. Moreover, accounting for the more than one million nursing days in all inpatient facilities in Hungary, this data with mostly negative patient experiences submitted in 10-year period could be considered small [35]. Another limitation is human interpretation used to evaluate models' performance as no quantitative metrics were found for the unsupervised models. Differentiation between the identified topics in the blog posts poses challenges, as a patient can mention different problems in one blog, and one problem could spread across sentences and paragraphs. Further research is needed to clarify methodological details such as the optional number of topics and differentiation between topics to provide more accurate results. Additionally, sentiment analyses were only applied to temporal categories, i.e., to blogs and comments divided by years, but not to the topics themselves. Further analysis is needed to assess the polarity of the identified topics and categories.

The Hungarian healthcare system provides health coverage for almost all its citizens based on universal health insurance. However, it is less exhaustive than in other EU countries [55], and 22.5% of the population reported unmet medical needs due to various difficulties (e.g., financial, waiting list, distance/transportation) in 2014 which is below the EU average. These factors can affect patient satisfaction and decision making, which could increase the possibility for complaints. Rules for investigating patient complaints are specified in the international regulation of healthcare providers that are also covered by national regulations. Patients can log complaints with the representatives in person, by phone, mail or email. According to the official website, studies evaluating patient satisfaction showed approximate satisfaction rates around 90% with 10% of negative feedback [36], in contrast to our findings. However, lack of information as to where and how to file an official complaint and/or lack of trust in official platforms prevented some patients from complaining formally. Several patients reported supposed uninterest from the representative of patient rights feeling that "they will not be taken seriously". While both providers and patients may fear consequences and detrimental effects [37], reporting complaints and responding to them online would help in a more direct, open communication, allowing providers to be better informed and implement necessary steps to improve their services [5].

The value of our research is that our findings potentially reveal the patterns and topics in patient experiences and directs attention on the examination of spontaneous non-solicited narratives, providing an opportunity to explore situations and encounters that probably cause the most frustration and dissatisfaction among patients. Most of the topics are related to limitations, shortcomings and deficits in health care services. The analysis of patient experiences – specially in large volume – could provide an important addition in improving the quality of care, especially if the analysis could be extended to names and localizations.

Our exploratory results showed the usefulness of BERT topic modeling to automatically process patient narratives. It highlights the need to develop an official patient online review web application embedded in Hungarian healthcare system to provide effective and continuous monitoring, which could contribute to the improvement of healthcare quality.

Conclusions

Our research aim was to automatically analyze and extract patterns of patient experiences and comments, made on the experiences of patients over a decade. Topic modeling, with BERT embeddings, were applied to identify frequently occurring topics and categories from both blog posts and comments. Five major categories were related to narratives, two of them – namely, structure and outcome category – were in consonant with previous theoretical frameworks for quality of care, and four categories emerge from the comments, respectively. Analysis was augmented by human interpretation to evaluate unsupervised models' performance, and to merge similar topics into categories. Sentiment analysis were used to categorized texts by polarity (positive, negative and neutral). The majority of texts were classified negatively by the model, implying dominantly negative health care encounters. To understand physician-patient encounter and problematic topics of healthcare services, NLP methods and techniques can be utilized effectively. However, the methods applied in our research, must be refined and improved, to strengthen the validity of our results and achieve less manual interpretation.

Funding

This research was funded by EFOP-3.6.1-16-2016-00022 "Debrecen Venture Catapult programme" project (co-financed by the European Union and the European Social Fund).

References

[1] A. Donabedian, "Evaluating the quality of medical care," Milbank Quarterly, Vol. 83, No. 4, Blackwell Publishing Inc., pp. 691-729, 2005, doi: 10.1111/j.1468-0009.2005.00397.x

[2] F. Carinci et al., "Towards actionable international comparisons of health system performance: Expert revision of the OECD framework and quality indicators," Int. J. Qual. Heal. Care, Vol. 27, No. 2, pp. 137-146, Apr. 2015, doi: 10.1093/intqhc/mzv004

[3] E. Larson, J. Sharma, M. A. Bohren, and Ö. Tunçalp, "When the patient is the expert: Measuring patient experience and satisfaction with care," Bull. World Health Organ., Vol. 97, No. 8, pp. 563-569, Aug. 2019, doi: 10.2471/BLT.18.225201

[4] Ó. Brito Fernandes et al., "Patient experiences with outpatient care in Hungary: results of an online population survey," Eur. J. Heal. Econ., Vol. 20, No. Suppl 1, pp. 79-90, Jun. 2019, doi: 10.1007/s10198-019-01064-z

[5] J. Li, M. Liu, X. Li, X. Liu, and J. Liu, "Developing embedded taxonomy and mining patients' interests from web-based physician reviews: Mixed-methods approach," J. Med. Internet Res., Vol. 20, No. 8, Aug. 2018, doi: 10.2196/jmir.8868

[6] K. Nawab, G. Ramsey, and R. Schreiber, "Natural Language Processing to Extract Meaningful Information from Patient Experience Feedback," Appl. Clin. Inform., Vol. 11, No. 2, pp. 242-252, 2020, doi: 10.1055/s-0040-1708049

[7] C. M. Burkle and M. T. Keegan, "Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians," BMC Health Serv. Res., Vol. 15, No. 1, pp. 1-7, Sep. 2015, doi: 10.1186/s12913-015-1099-2

[8] V. Zsuzsa and H. Tamás, "Patients' preferences for health-related use of Internet" Orv. Hetil., Vol. 159, No. 51, pp. 2175-2182, 2018, doi: 10.1556/650.2018.31210

[9] T. Lagu, S. L. Goff, N. S. Hannon, A. Shatz, and P. K. Lindenauer, "A mixed-methods analysis of patient reviews of hospital care in England: Implications for public reporting of health care quality data in the United States," Joint Commission Journal on Quality and Patient Safety, Vol. 39, No. 1, Joint Commission Resources, Inc., pp. 7-15, Jan. 01, 2013, doi: 10.1016/s1553-7250(13)39003-5

[10] R. A. Price et al., "Examining the role of patient experience surveys in measuring health care quality," Medical Care Research and Review, Vol. 71, No. 5, SAGE Publications Inc., pp. 522-554, Oct. 08, 2014, doi: 10.1177/1077558714541480

[11] A. S. Gallan, M. Girju, and R. Girju, "Perfect ratings with negative comments: Learning from contradictory patient survey responses," Patient Exp. J., Vol. 4, No. 3, pp. 15-28, 2017, doi: 10.35680/2372-0247.1234

[12] M. Khanbhai, P. Anyadi, J. Symons, K. Flott, A. Darzi, and E. Mayer, "Applying natural language processing and machine learning techniques to patient experience feedback: A systematic review," BMJ Health and Care Informatics, Vol. 28, No. 1, BMJ Publishing Group, p. 100262, Mar. 02, 2021, doi: 10.1136/bmjhci-2020-100262

[13] A. López, A. Detz, N. Ratanawongsa, and U. Sarkar, "What patients say about their doctors online: A qualitative content analysis," J. Gen. Intern. Med., Vol. 27, No. 6, pp. 685-692, 2012, doi: 10.1007/s11606-011-1958-4

[14] A. Le Glaz et al., "Machine Learning and Natural Language Processing in Mental Health: Systematic Review," J. Med. Internet Res., Vol. 23, No. 5, May 2021, doi: 10.2196/15708

[15] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson, "Use of sentiment analysis for capturing patient experience from free-text comments posted online," J. Med. Internet Res., Vol. 15, No. 11, pp. 1-9, 2013, doi: 10.2196/jmir.2721

[16] K. Doing-Harris, D. L. Mowery, C. Daniels, W. W. Chapman, and M. Conway, "Understanding patient satisfaction with received healthcare

services: A natural language processing approach," AMIA ... Annu. Symp. proceedings. AMIA Symp., Vol. 2016, pp. 524-533, Jan. 2016, Accessed: Nov. 17, 2020 [Online] Available: https://europepmc.org/articles/PMC5333198

[17] D. Angelov, "Top2Vec: Distributed Representations of Topics," arXiv, Aug. 2020, Accessed: Apr. 27, 2021 [Online] Available: http://arxiv.org/abs/2008.09470

[18] B. Saha, S. Lisboa, and S. Ghosh, "Understanding patient complaint characteristics using contextual clinical BERT embeddings," 2020

[19] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., Vol. 1, No. Mlm, pp. 4171-4186, 2019

[20] D. M. Nemeskey, "Natural Language Processing Methods for Language Modeling", Doctoral Dissertation, Doctoral School of Informatics, Faculty of Informatics, Eötvös Lóránd University, 2020

[21] D. M. Nemeskey, "Introducing huBERT," in XVII. Conference on Hungarian Computational Linguistics, 2019, pp. 3-14

[22] J. Ács, D. Lévai, D. M. Nemeskey, and A. Kornai, "Evaluating Contextualized Language Models for Hungarian," Feb. 2021, Accessed: Apr. 27, 2021 [Online] Available: http://arxiv.org/abs/2102.10848

[23] "HIL-SBERT." https://hilanco.github.io/models/sbert.html (accessed Jan. 20, 2022)

[24] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf., pp. 3982-3992, Aug. 2019, Accessed: Jul. 01, 2021 [Online] Available: http://arxiv.org/abs/1908.10084

[25] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," Feb. 2018, Accessed: Jul. 01, 2021 [Online] Available: http://arxiv.org/abs/1802.03426

[26] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," J. Open Source Softw., Vol. 2, No. 11, p. 205, 2017, doi: 10.21105/joss.00205

[27] "MaartenGr/cTFIDF: Creating class-based TF-IDF matrices." https://github.com/MaartenGr/cTFIDF (accessed Jan. 20, 2022)

[28] "Hungarian Twitter Sentiment Corpus - opendata.hu." http://opendata.hu/dataset/hungarian-twitter-sentiment-corpus (accessed Jan. 20, 2022)

[29] "Precognox." https://www.precognox.com/ (accessed Jan. 20, 2022)

[30] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding," 7th Int. Conf. Learn. Represent. ICLR 2019, Apr. 2018, Accessed: Jul. 01, 2021 [Online] Available: http://arxiv.org/abs/1804.07461

[31] Ó. B. Fernandes et al., "Self-reported waiting times for outpatient health care services in Hungary: Results of a cross-sectional survey on a national representative sample," Int. J. Environ. Res. Public Health, Vol. 18, No. 5, pp. 1-15, Mar. 2021, doi: 10.3390/ijerph18052213

[32] A. J. Forster, H. J. Murff, J. F. Peterson, T. K. Gandhi, and D. W. Bates, "The incidence and severity of adverse events affecting patients after discharge from the hospital," Ann. Intern. Med., Vol. 138, No. 3, pp. 161-167, Feb. 2003, doi: 10.7326/0003-4819-138-3-200302040-00007

[33] M. Bahja and M. Lycett, "Identifying patient experience from online resources via sentiment analysis and topic modelling," Proc. - 3rd IEEE/ACM Int. Conf. Big Data Comput. Appl. Technol. BDCAT 2016, pp. 94-99, 2016, doi: 10.1145/3006299.3006335

[34] T. A. Ito, J. T. Larsen, N. K. Smith, and J. T. Cacioppo, "Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations," J. Pers. Soc. Psychol., Vol. 75, No. 4, pp. 887-900, 1998, doi: 10.1037//0022-3514.75.4.887

[35] "National Health Insurance Fund of Hungary – Inpatient Statistics." http://www.neak.gov.hu/virtualis_rovat/altfin_virt_dok2/besorolo/fekvo_stat (accessed Nov. 21, 2021)

[36] "Investigating Patients' Complaints." https://www.patientsrights.hu/investigating-patients-complaints.html (accessed May 31, 2021)

[37] T. Bourne et al., "Doctors' experiences and their perception of the most stressful aspects of complaints processes in the UK: An analysis of qualitative survey data," BMJ Open, Vol. 6, No. 7, pp. 1-10, 2016, doi: 10.1136/bmjopen-2016-011711

[38] Creative Commons Hungary, https://creativecommons.org/licenses/by-nc-nd/2.5/hu/ (accessed Jan 25, 2022)