A framework of attribute extraction and dependable aspect term selection from reviews of hospital websites

Nasreen Taj Mohammed Basha[1], Girisha Gowdra Shivappa[2]
[1]Department of Information Science and Engineering, Gowdara Mallikarjunappa Institute of Technology, Davanagere, India
[2]Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru, India

Article Info

Article history:

Received Dec 5, 2023
Revised Feb 23, 2024
Accepted Mar 12, 2024

Keywords:

Aspect term extraction
Attribute extraction
Coherence-score
Hospital
Reviews
Topic modelling

ABSTRACT

Online reviews found on hospital websites and external platforms constitute user-generated content where patients and their families share their firsthand encounters. As patients increasingly rely on online platforms to share their experiences, understanding the importance of their feedback is paramount for healthcare providers. The novelty of this research lies in the development of advanced frameworks that not only extract relevant information but also offer a more sophisticated and coherent analysis of the multifaceted aspects embedded in patient reviews. Hence, this work involves collecting data from various hospital websites, followed by data pre-processing to ensure accuracy and consistency. Subsequently, two distinct frameworks are proposed. The first framework aims to extract specific attributes (topics) mentioned in reviews, enhancing the granularity of information derived from the collected data. The second framework addresses the efficient extraction of aspect terms from pre-processed data, utilizing a coherence score-based approach called as modified latent dirichlet allocation term frequency-inverse document frequency (M-LDA TF-IDF). The M-LDA TF-IDF has achieved better a coherence score of 0.478 which is much better in comparison with other topic modelling approaches.

Corresponding Author:

Nasreen Taj Mohammed Basha
Department of Information Science and Engineering, Gowdara Mallikarjunappa Institute of Technology
P.B No. 577006, 4th Main Rd, Davanagere, Karnataka, India
Email: nasreent@gmit.ac.in

1. INTRODUCTION

In today's world, reviews play a crucial role in aiding people's understanding of products, services, and experiences [1]. Online reviews found on hospital websites and external platforms constitute user-generated content where patients and their families share their firsthand encounters and viewpoints regarding the care, services, and amenities received during their hospital stays [2]. These reviews, accessible on hospital websites as well as third-party review sites and social media [3], aim to furnish potential patients with valuable insights into the hospital's care quality and overall patient involvement. The reviews aid patients in informed decision-making by assessing care quality, evaluating the patient experience encompassing facility hygiene, room comfort, and staff responsiveness, offering insights into location convenience and accessibility, providing relatable experiences for patients with comparable medical conditions, enabling hospital comparisons based on diverse factors, and enabling hospitals to utilize feedback to enhance services [4].

The increased reliance on online reviews for hospital appointments post-COVID-19 arises from multiple factors [5]. Firstly, heightened safety concerns due to the pandemic have led individuals to prioritize healthcare facilities with severe safety measures, and online reviews offer insights into hospitals' safety protocols and infection control practices [6]. Additionally, the shift to online activities during lockdowns has prompted people to assess hospitals remotely, avoiding in-person visits and traditional word-of-mouth recommendations [7]. Online reviews provide a means to gauge patient experiences, enabling potential patients to make informed decisions based on insights into care quality, staff communication, wait times, and overall satisfaction [8]. Moreover, reviews fill information gaps left by hospital websites, offering a transparent and comprehensive view of services and patient experiences [9]. The accessibility and convenience of online reviews are crucial, particularly for those with mobility challenges or health conditions [10]. Furthermore, positive experiences shared in reviews can alleviate anxiety associated with medical visits [11]. These reviews also contribute to quality assurance, building confidence in hospitals ability to provide effective care and a positive experience [12]. The sense of community support highlighted during the pandemic encourages individuals to back hospitals with positive feedback [13]. Lastly, the continuation of digital behaviors, spurred by the pandemic, has led to ongoing reliance on online research and appointment scheduling [14].

However, efficiently gathering data from hospital websites presents challenges due to website diversity, technical complexity, and data protection concerns [15], [16]. Some strategies to address these challenges involve custom-made scraping solutions, application programming interface (API) integration, vigilant monitoring, ethical compliance, proxy usage, and data validation [17]. Successful data collection demands technical proficiency, adaptability, and adherence to legal and ethical standards, acknowledging the dynamic and multifaceted landscape of hospital websites [18]. Hence, in this work, we have collected data from 200+ hospitals from the mouthshut.com where multiple patients have given reviews for either doctor, staff or hospital's overall experience. A total of 50000+ individual reviews were collected from these hospitals. The reviews also consisted of emojis which have been removed during pre-processing. The data was further pre-processed and a framework was built for the extraction of specific attributes which will help to analyze the sentiment of the individual. The contribution of this work is as follows: i) collect data from various hospitals websites, ii) perform data pre-processing technique to clean the collected data, iii) design a framework for extraction of specific attributes mentioned in the reviews from the pre-processed data, and iv) design a framework which will help to efficiently extract aspect terms from the preprocessed data.

2. LITERATURE SURVEY

In this literature survey, various methods used for the collection of data and pre-processing techniques used for aspect term extraction (ATE) has been given. Using a variety of pre-trained embeddings of words, Augustyniak *et al.* [18] zeroed down on long short-term memory (LSTM)-based basic frameworks featuring the possibility of conditionally-randomized-fields (CRF) for better improvement for ATE. The experiments carried out on SemEval datasets demonstrated that bi-directional long short-term memory (BiLSTM) can be utilized as an exceptional predictor, even when compared to extremely advanced and intricate algorithms employing enormous word embeddings or linguistic models. For the purpose of performing aspect based-sentiment analysis (ABSA), Huang *et al.* [19] presented an innovative logic-tensor-network utilizing massive rules (LTNMR) that is built using the aid of first-order logic (FOL). They introduced a mutual distillation structural knowledge injection (MDSKI) technique to improve inferential precision. Using MDSKI, BERT, the teacher, imparted his knowledge of dependencies to the LTNMR. Both SemEval-14 along with the Twitter dataset were used. For ABSA, experimental results show that the suggested LTNMR, when paired with the standard MDSKI method, substantially surpasses the latest findings of existing works.

Multi-domain keyword-extraction utilizing word vectors is a technique proposed in [20] through the goal of improving an individual's satisfaction by providing them with comprehensive reviews coming from numerous websites along with comprehensive evaluations of those reviews. A special approach was designed to find reliable reviews, and it took into account a number of factors that establish the reliability of a review. The outcomes of the trials conducted on real-time data sets were superior to those of the previously used traditional approaches. A unique Seq2Seq framework, the information-augmented neural-network (IANN), was proposed in [21]. In IANN, the encoder was built around a particular neural-network called a multiplexed-convolutional alongside recurrent-network (MCRN), which is used for encoding the most relevant data and incorporate contextual data from neighboring words. In order to accurately represent the ever-changing meaning of words, an integrated embedding structure has been developed. These findings confirm the studies that suggested IANN is an effective technique for the ATE tasks and show that it is able to achieve better state-of-the-art outcomes.

By drawing on the experience and expertise of earlier trained computational language approaches, [22] proposed a transfer learning (TL)-based method for ATE and aspect polarity detection (APD) in Arabic. The proposed approach was built upon the BERT model's Arabic base (Arabic version). There is additionally a comparison of several BERT algorithms. The studies employed the HAAD dataset, which is based on the ABSA standard dataset. The experimental findings show that their approach was superior to both the reference approach and the existing approaches. Multiple feature-vectors are extracted simultaneously using dilated convolution layers in [23], and these are subsequently combined for classification. Using reinforcement method, where the training approach was portrayed as a consecutive process of decision-making, they were able to reclaim the ATE approach from imbalanced categorization. When tested on two English-language datasets (laptop and restaurant), the suggested approach outperformed state-of-the-art deep-learning approaches with impressive precision and F-measure values. The model achieved 85.44 percent and 87.35 percent of precision and F-measure for restaurant dataset and 80.88 percent and 80.78 percent of precision and F-measure for laptop dataset. Bansal and Kumar [24], used sentiment evaluation to decipher the plethora of internet hospital reviews available to help patients make informed decisions. Over 500 different hospitals contributed their feedback, totaling in excess of 30,000 reviews. Overall, this research aimed to provide patients with an accurate and detailed ranking of hospitals according to the opinions of actual patients who have used those facilities.

The literature survey highlights several notable research gaps in the field of sentiment analysis and review classification. Firstly, previous research predominantly concentrated on the utilization of the SemEval dataset for evaluation purposes, with limited exploration into other datasets such as the hospital review dataset, as evidenced by only one study [24]. Secondly, a common shortcoming among existing studies was inadequate data preprocessing, notably the omission of emoji removal from review texts. Furthermore, the study [24] manually selected attributes, suggesting the potential for more automated or data-driven attribute selection methods in future research. Moreover, while [24] focused on just four aspects, the proposed model encompasses a more comprehensive consideration of six different aspects, indicating a broader scope of analysis. Lastly, existing studies primarily operated on small datasets from a single source, whereas the proposed model deals with a substantial dataset collected from multiple sources, expanding the scale and diversity of the data under examination.

3. METHOD

The framework for the proposed work is presented in Figure 1. In this presented framework, the collected data from 200+ hospitals are preprocessed. In the first step of preprocessing, the NULL values are dropped from the dataset. In the second step, the complete dataset is converted into lower case and all the non-alphabetical data such as punctuations and numbers are removed from the dataset. Word-tokenization occurs in the next stage. Tokenizing words is a method of breaking down a huge text samples into individual words. In natural language processing (NLP) tasks, such as sentiment evaluation and counting words, it is necessary to record every word in order to analyze them individually. To do this, this work utilized the NLTK library, which is devoted to processing natural-language. In the fourth step, the unnecessary words are removed using the stopwords library. In the fifth step, the lemmatization is done. Combining collectively various variations of a particular word that have been altered is called lemmatization. Finally, in the final step, that is, sixth step, part of speech (PoS) tagging is done. POST or PoS tagging also known as grammar tagging, is a method of assigning a word inside a text (corpus) upon a specific PoS according to how that word is typically used and how it is defined. After the PoS tagging, the preprocessed goes through n-grams, modified latent dirichlet allocation (M-LDA) and then finally the ATE takes place.

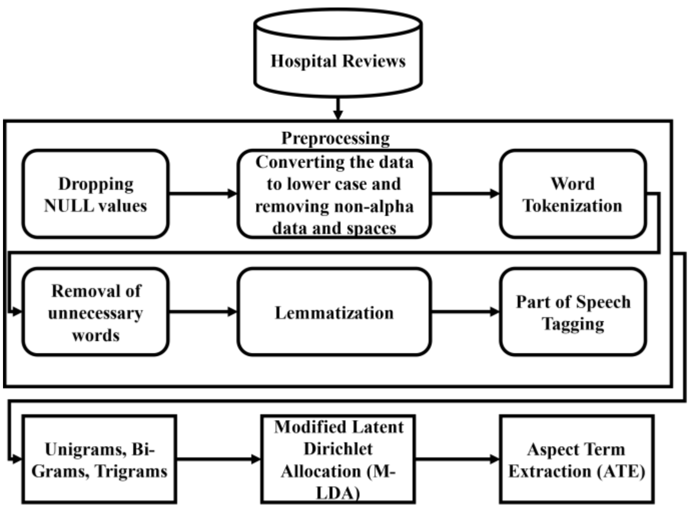

Figure 1. Framework for ATE

3.1. Part of speech tagging

In this work, we have preprocessed the data and extracted the aspects using the stochastic supervised taggers. In the Figure 2, the complete process of the PoS tagging has been shown. The hospital preprocessed reviews first go through the sentence transformation and word segmentation. Further, the PoS tags are constructed using shallow parser tool. Using the hidden markov model (HMM), the transition probability of each review is evaluated. The lexical and syntactic similarities between the reviews are exploited.

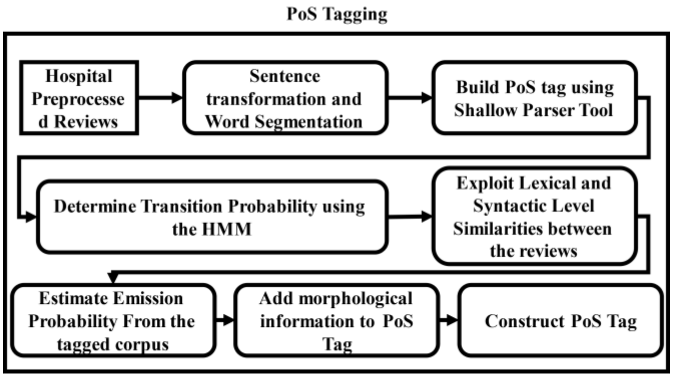

Figure 2. PoS tagging

3.2. N-gram

In this work, we employ a second-order Markov method to n-grams tags to create a PoS tagger for the hospital reviews. TnT [25] applied to a POS-tagger is just as effective as any different approach, as demonstrated in [26]. The likelihood of both transitions and emissions make up the tagger framework. Tags are represented by the framework's states/transitions, whereas words are represented by their emissions. The probability of changing between states are proportional to the number of pairings of tags. The likelihood of an emission occurring is mostly determined by the newest or most latest category. Using the (1), tag sequences for a particular word are selected.

$$\underset{\mathscr{k}\ldots\mathscr{k}_{\mathcal{K}}}{\operatorname{argmax}}\left[\prod_{m=1}^{\mathcal{K}} \mathcal{P}(\mathscr{k}_m|\mathscr{k}_{m-1},\mathscr{k}_{m-2})P(\mathscr{j}_m|\mathscr{k}_i)\right] P(\mathscr{k}_{\mathcal{K}+1}|\mathscr{k}_{\mathcal{K}}) \tag{1}$$

Where, PoS tags are denoted by $\mathscr{k}_m \ldots \mathscr{k}_{\mathcal{K}}$, supplementary/extra tags are denoted by $\mathscr{k}_{-1}$, $\mathscr{k}_0$, and $\mathscr{k}_{\mathcal{K}+1}$. The $\mathscr{k}_{-1}$ represents beginning of sequenced marker, $\mathscr{k}_0$ represents middle of sequenced marker and $\mathscr{k}_{\mathcal{K}+1}$ represents ending of sequenced marker. Further, word sequences are denoted as $\mathscr{j}_m \ldots \mathscr{j}_{\mathcal{K}}$ for the word-length $\mathcal{K}$. The existing frameworks presented in the recent year ends on a "loose-end" at the lost terms, however by using these supplemental tags, we can improve the tagging outcomes. Nevertheless, if no sentence's limits are defined, the proposed method tagger will insert them if it recognizes any of the tokens [.,?,!, ;]. The likelihood of a transition or an emission can be calculated using a labeled data set. The maximum-likelihood $\hat{\mathcal{P}}$ are first derived from relative frequencies.

$$Unigrams: \hat{\mathcal{P}}(\mathscr{k}_3) = {\mathcal{R}(\mathscr{k}_3)}/{\mathcal{T}} \tag{2}$$

$$Bigrams: \hat{\mathcal{P}}(\mathscr{k}_3|\mathscr{k}_2) = {\mathcal{R}(\mathscr{k}_2,\mathscr{k}_3)}/{\mathcal{R}(\mathscr{k}_2)} \tag{3}$$

$$Trigrams: \hat{\mathcal{P}}(\mathscr{k}_3|\mathscr{k}_1,\mathscr{k}_2) = {\mathcal{R}(\mathscr{k}_1,\mathscr{k}_2,\mathscr{k}_3)}/{\mathcal{R}(\mathscr{k}_1,\mathscr{k}_2)} \tag{4}$$

$$Lexical: \hat{\mathcal{P}}(\mathscr{j}_3|\mathscr{j}_2) = {\mathcal{R}(\mathscr{j}_3,\mathscr{k}_3)}/{\mathcal{R}(\mathscr{k}_3)} \tag{5}$$

For all the word-sequence $\mathscr{j}_3$ inside lexicons, as well as the PoS tags $\mathscr{k}_1,\mathscr{k}_2,\mathscr{k}_3$ within a tag set, $\mathcal{T}$ is the total number of tokens within the set of data used for training. In this study, it is assumed or stated the fact that ML likelihood is also zero if the denominator and numerator of the matrix are both zero. In addition, out-of-lexicon terms are used to standardize contextually frequencies, and lexicon frequencies are treated in the same way. In most cases, the lack of information issue prevents direct use of trigram likelihood construction from dataset. Therefore, there is insufficient data for every trigram to confidently calculate the likelihood. On the other hand, if the related trigram did not exist throughout the dataset, setting the likelihood to 0 would have an unintended result. Since it is difficult to rank distinct sequencing with a zero likelihood, the likelihood of a whole sequence becomes zero if its employment is required for an entirely novel sequence. Linear interpolation of $trigrams$, $bigrams$, and $unigrams$ is the optimal normalizing variable in $Trigrams'n'Tags$. Following is the (6) for determining the likelihood of a trigram.

$$\hat{\mathcal{P}}(\mathscr{k}_3|\mathscr{k}_1,\mathscr{k}_2) = \beta_1\hat{\mathcal{P}}(\mathscr{k}_3) + \beta_2\hat{\mathcal{P}}(\mathscr{k}_3|t_2) + \beta_3\hat{\mathcal{P}}(\mathscr{k}_3|\mathscr{k}_1,\mathscr{k}_2) \tag{6}$$

Where, $\mathcal{P}$ denotes the likelihood distributions. Nevertheless, as $\hat{\mathcal{P}}$ is obtained using the likelihood approximation, and $\beta_1 + \beta_2 + \beta_3 = 1$. Because we are employing a context-free method of interpolation using linearity, the numerical value of βs in presented work does not rely on the specific trigram. This helps get better results than the current best context-dependent method. Given the lack of information issue, it is not possible to calculate an individual collection of βs for every trigram. As a result, we calculate collections of βs that are grouped according to trigram frequency. To the best of our knowledge, no earlier research has examined frequency groupings for interpolation using linearity in tagging with POS. Using removed interpolation, we can calculate $\_\beta_1$, β_2, and β_3 values. This method aids in computing ideal/optimal settings for each βs by eliminating every trigram coming from the experimental dataset one by one. In a computational time, proportional to the number of distinct trigrams, the weights can be efficiently determined by first finding the counts of frequencies for $trigrams$, $bigrams$, and $unigrams$. The complete process of how the preprocessing and aspect extraction is done which is shown as follows.

Pre-processing and filtering aspects:

– Step 1: begin with a collection of reviews denoted as D={d_1,d_2,d_3,…d_m } describing various aspects within the healthcare domain C={c_1,c_2,c_3,….c_m } with n>1 and m>1. Our objective in this endeavor is to pinpoint and categorize the most pertinent aspects from C. In order to initiate this process, the reviews d_(i)∈D are subjected to tokenization, thereby generating n-grams that serve as initial candidate aspects c_(i)∈C. We perform this operation for all d_(i)∈D.
– Step 2: in these n-grams, we proceed to eliminate stop words utilizing a list containing common terms. Furthermore, we calculate the frequency of occurrence of the remaining aspects c_i in D, resulting in a collection of tuples represented as s = {(c_i, f_i.),….}. Each tuple pairs an aspect c_i with its corresponding frequency f_i.
– Step 3: to mitigate noise, we execute the removal of short and infrequent n-grams from T, and by extension, from C. This is accomplished by applying a frequency threshold denoted as f_t.
– Step 4: Solely unigrams (with n>=1) that surpass the frequency threshold f_t and occur notably more frequently than any larger n-gram containing the same unigram within its structure are retained as meaningful aspects.
– Step 5: if there exists more than one n-gram c_j of higher order encompassing c_i, f_i is associated with the most frequently transpiring c_j. The remaining n-grams from C are merged according to two specific rules to ensure the presence of only a singular n-gram for semantically related n-grams. Firstly, any plural token is excluded if its singular form is also present. Secondly, the present participle of a regular verb is discarded if an equivalent form without it exists.
– Step 6: among the aspects left within C, those lacking a corresponding DBpedia entries are filtered out.
– Step 7: we proceed to construct initial facet information denoted as C_info and categorize d_i based on distinct terms such as doctor, service, and staff.

3.3. Aspect term extraction

The term frequency-inverse document frequency (TF-IDF) model is an essential tool in text analysis, particularly in the context of latent-dirichlet-allocation (LDA) [27]. LDA, a prevalent machine-learning clustering method, facilitates the extraction of aspects from textual data. In Figure 3, the M-LDA TF-IDF process is illustrated, highlighting the intricate flow of the proposed model. The TF-IDF model plays a crucial role in LDA's first phase, where a corpus of reviews is generated. It assists in quantifying the importance of each term in a document within the context of the entire corpus, contributing to the identification of relevant topics and aspects. In phase 1, LDA assumes a generative process for creating reviews, where the TF-IDF model aids in drawing distributions over topics for each document. TF-IDF helps assign weights to terms, emphasizing their significance in capturing the essence of the topics. During inference (phase 2), TF-IDF continues to play a role in estimating the latent variables by evaluating the probability distributions of the observed data, ensuring a comprehensive understanding of the hidden structures within the corpus. In phase 3, TF-IDF contributes to the evaluation of coherence scores (CS) for each review. The CS, assessed using coherence validation (CV) metrics, relies on TF-IDF to gauge the similarity between words associated with specific topics. The TF-IDF model, with its emphasis on term importance, aids in generating content vectors and computing cosine similarity measures, essential for determining the coherence and interpretability of the identified topics. Overall, TF-IDF serves as a critical component in enhancing the effectiveness and interpretability of LDA in the context of topic modeling. The phases are given as follows.

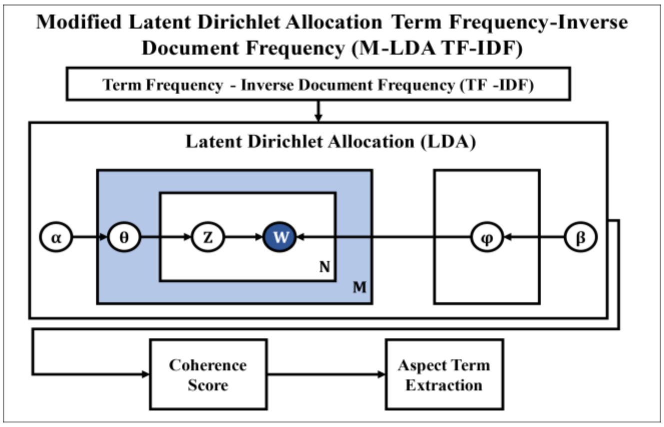

Figure 3. Latent dirichlet allocation

3.3.1. Phase 1

LDA assumes that a corpus (collection) of reviews is generated based on a hidden structure of topics. Each document is a mixture of various topics, and each topic is a distribution over words (aspects). The model aims to uncover these latent topics and their associated aspects distributions. LDA assumes a generative process for creating a corpus of reviews:

a. For each review i (document) in the corpus:
- Draw a distribution over topics for the reviews from a Dirichlet distribution [28] with parameter X. This distribution represents the mixture of topics in the document.
- This step generates the topic proportions for the document.

b. For each word T in the review i:
- Draw a topic j from the previously drawn distribution over topics for the document i.
- This step assigns a topic to each aspect in the review.

c. Draw a word W from the topic's distribution over words, which is defined by a Dirichlet distribution parameter Y.
- This step selects a specific word from the chosen topic's word distribution.

3.3.2. Phase 2

In this phase the inference is done. The goal of inference in LDA is to reverse-engineer the process that generated the corpus. Given a collection of reviews, we want to determine the hidden variables (topic mixtures for each review and aspect assignments to topics) that likely produced these reviews. Inference is often performed using techniques like Gibbs sampling [29] or variational inference [30]. These methods estimate the probability distributions of the latent variables based on the observed data (the words in the reviews).

3.3.3. Phase 3

a. Topic distributions for reviews:
- After performing inference, each document is represented as a distribution over topics. The weights in this distribution indicate the strength of each topic's presence in the review.
- This output provides insight into which topics are prominent in each review.

b. Word distributions for topics:
- Similarly, after inference, each topic is represented as a distribution over words. The weights in this distribution indicate the likelihood of each word appearing in the topic.
- This output helps in interpreting the aspects associated with each topic.

Given a corpus of reviews, the goal of LDA is to infer the latent variables (topic mixtures for each document and word (aspects) assignments to topics) that generated the corpus. Using all these parameters, the CS for each review is evaluated. The CS, which is employed in topic-modeling, serves as a metric for evaluating the extent to which the generated topics can be comprehended and interpreted by individuals. In this particular scenario, the topics under consideration are denoted by the highest-ranking N words exhibiting the greatest likelihood of being associated through the specific topic in question. The CS is a metric that quantifies the degree of similarity between the words under consideration. One of the finest and most widely utilized metrics for assessing coherence in various domains is commonly referred to as CV. The process involves generating content vectors by analyzing the co-occurrence patterns of words. Subsequently, a scoring mechanism is employed, utilizing cosine similarity measures along with normalized pointwise mutual information (NPMI). The popularity of this metric can be attributed to its utilization as the default metric within the Gensim topic coherence pipelines component. The CV coherence is evaluated using the (7) [31].

$$\phi S_i(\vec{u},\vec{w}) = \frac{\sum_{i=1}^{|W|} u_i \cdot w_i}{||\vec{u}||_2 \cdot ||\vec{w}||_2} \tag{7}$$

The results achieved by the following work have been discussed in the results and discussion section.

4. RESULTS AND DISCUSSION

The results have been evaluated in this section. The system used for evaluation consisted i7 processor, 16 GB RAM, and Windows 11 operating system. The coding was done in Python. For evaluation of this work, the hospital dataset was considered. The most talked topics in the dataset were analyzed and it was found that 6 topics were the most repeated in the reviews. The 6 topics include hospital, patient, surgery, doctor, staff, and service. The results achieved for the hospital dataset is presented in Figure 4.

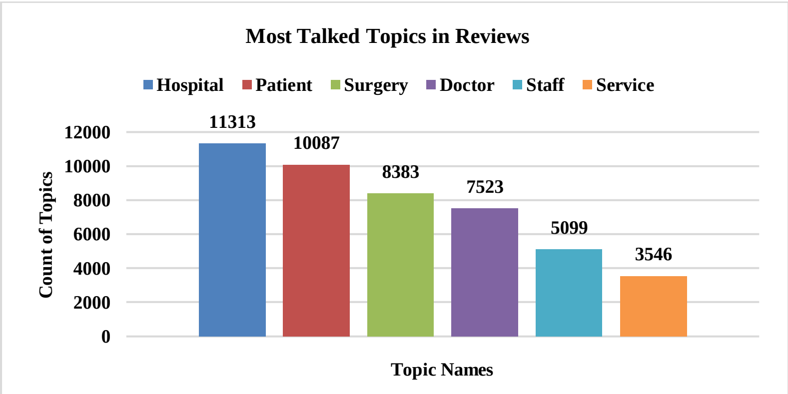

Figure 4. Most talked topics in reviews

Further, after identification of the most repeated topics, the results have been accumulated in Tables 1 to 3. In Table 1, the review classification for the topics is presented where each review represents the following aspect. Further, aspects were extracted from each review. Some of the examples are discussed in Table 2. Finally, in this work, the different aspects representing a given topic is presented using Table 3.

This work has compared the CV scores with different topic modelling approaches such as hierarchal dirichlet process (HDP), latent semantic analysis (LSA), and LDA. A higher CS generally indicates better interpretability and cohesion among identified topics. In this comparison, presented in Figure 5, M-LDA stands out with a significantly higher CS of 0.4784, showcasing its superior ability to generate coherent and contextually relevant topics. LDA follows closely with a CS of 0.3650, demonstrating its effectiveness in extracting meaningful topics from the given corpus. LSA, with a CS of 0.2914, exhibits a relatively lower coherence, suggesting that it may not capture the underlying thematic structures as effectively as the other models. HDP, with a CS of 0.1783, trails behind, indicating a potential limitation in coherence compared to the M-LDA and LDA models. These coherence scores provide a quantitative basis for evaluating the performance of different topic modelling approaches, guiding the selection of the most suitable algorithm based on the desired interpretability and cohesion of identified topics.

Table 1. Review classification according to topics

Review	Topic
I have been visiting Dentzz for the last two years and can vouch for them. The service is premium and you can put all your worries aside and trust the doctor's opinion. They are very professional yet accommodating of your needs. Best dentistry around.	Doctor
I was amazed to see procedure with warranty here, I got my fillings with 3-year warranty! Apart from that, the process was smooth and did not trouble my schedule at all. Their services are quick, especially their technical part of work is updated and causes no pain during treatments. Love the work at Dentzz.	Surgery
Good service with excellent doctor, the environment is good, well-maintained hygiene. Overall services are good.	Service
Their private patient care is best. Doctor s are available at 8: 30 am itself. Especially uro doctors. They will be available in opd @8 am itself.	Patient
Good hospital doctors and staffs are very polite they made me comfort good infrastructure, very friendly, Im impressed with the hospitality	Staff
I have always given a lot of importance to dental and oral healthcare and make sure that I visit my dentist at Dentzz, once every 2 months.	Hospital

Table 2. Topics having aspects

Review	Aspects
I have been visiting Dentzz for the last two years and can vouch for them. The service is premium and you can put all your worries aside and trust the doctor's opinion. They are very professional yet accommodating of your needs. Best dentistry around.	Visiting, Doctor, Service,
I was amazed to see procedure with warranty here, I got my fillings with 3-year warranty! Apart from that, the process was smooth and did not trouble my schedule at all. Their services are quick, especially their technical part of work is updated and causes no pain during treatments. Love the work at Dentzz.	Procedure, See, Warranty, Year, Process, Trouble, Schedule, Part, Work, Dentzz
Good service with excellent doctor, the environment is good, well-maintained hygiene. Overall services is good.	Good, Service, Doctor, Environment, Hygiene
Their private patient care is best. Doctor s are available at 8: 30 am itself. Especially uro doctors. They will be available in opd @8 am itself.	Patient Care, Private, OPD
Good hospital doctors and staffs are very polite they made me comfort good infrastructure, very friendly, Im impressed with the hospitality	Hospital, Comfort, Good, Infrastructure, Hospitality
I have always given a lot of importance to dental and oral healthcare and make sure that I visit my dentist at Dentzz, once every 2 months.	Lot, Importance, Healthcare, Visit, Dentist, Dentzz

Table 3. Topics having aspects

Topic	Aspect and its Attributes
Doctor	Visiting, Doctor, Service,
Surgery	Procedure, See, Warranty, Year, Process, Trouble, Schedule, Part, Work, Dentzz
Service	Good, Service, Doctor, Environment, Hygiene
Patient	Patient Care, Private, OPD
Staff	Hospital, Comfort, Good, Infrastructure, Hospitality
Hospital	Lot, Importance, Healthcare, Visit, Dentist, Dentzz

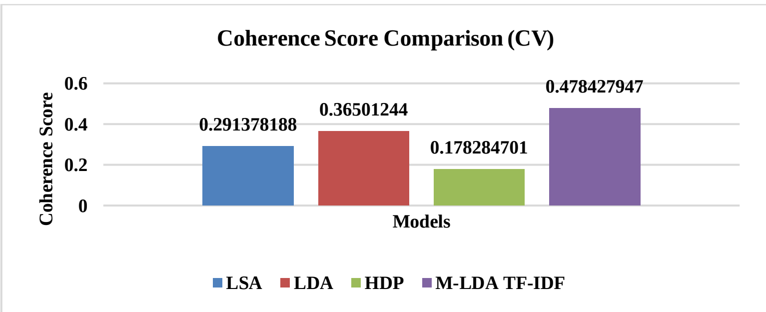

Figure 5. Coherence scores comparison

The Table 4 presents the performance metrics of various topic modeling algorithms, LSA [32], LDA [32], HDP [32], and M-LDA TF-IDF, across different datasets. Looking at the results, it iss evident that the M-LDA TF-IDF algorithm consistently outperforms the other three methods across all datasets. For instance, in the mobile dataset, M-LDA TF-IDF achieves a score of 0.487212, surpassing the scores of LSA (0.28307), LDA (0.41770), and HDP (0.19771). This trend is observed consistently in other datasets like IMDB, musical instruments, automotive, restaurant, and hospital. The superiority of M-LDA TF-IDF is attributed to its ability to incorporate both topic modeling (LDA) and term weighting (TF-IDF), leveraging the strengths of both techniques. This helps to provide a better balance in capturing latent topics within the datasets, resulting in more accurate and meaningful representations of the underlying structures. On the other hand, LSA and LDA exhibit competitive performances, with LDA outperforming LSA in most cases. HDP consistently demonstrates the weakest performance among the four algorithms, suggesting that the hierarchical approach may not be as effective in capturing latent topics compared to the other methods. In summary, the M-LDA TF-IDF algorithm appears to be the most effective choice for topic modeling in this context, offering superior performance across diverse datasets compared to traditional LSA, LDA, and HDP methods.

Table 4. Comparative study

Dataset	LSA [32]	LDA [32]	HDP [32]	M-LDA TF-IDF
Mobile	0.28307	0.41770	0.19771	0.487212
IMDB	0.31487	0.357489	0.23234	0.568341
Musical Instruments	0.34856	0.36402	0.33599	0.483012
Automotive	0.37685	0.45847	0.34492	0.547893
Restaurant	0.40988	0.46909	0.375695	0.572462
Hospital	0.29137	0.36501	0.17828	0.478427

5. CONCLUSION

In the work, the initial step involved data preprocessing, where the collected data underwent a specific technique known as PoS tagging, along with the application of n-gram analysis. PoS tagging involved assigning grammatical labels to each word in a text, categorizing them based on their roles within sentences, such as nouns, verbs, adjectives, and so on. N-grams refered to contiguous sequences of n words from a given text, and their analysis provides insight into patterns and relationships between words. After this, the work focused on ATE where, the M-LDA TF-IDF was presented which comprised of LDA and TF-IDF. The extraction of these aspect terms was carried out using a M-LDA TF-IDF that involved assessing the coherence score. The results showed that proposed M-LDA TF-IDF achieved higher coherence score in comparison with LSA, LDA, and HDP. Also, this work helped to identify aspect terms by recognizing words that often appear together in a meaningful manner. In future work, this work would be further extended to understand the sentiments from the reviews.

REFERENCES

[1] M. Kang, B. Sun, T. Liang, and H. Y. Mao, "A study on the influence of online reviews of new products on consumers' purchase decisions: an empirical study on JD.com," *Frontiers in Psychology*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.983060.

[2] S. K. S. S. K.S., P. Supriya, G. Barkur, S. G. Somu G., and N. Sampathila, "Analysis of hospital website contents for promoting medical tourism among accredited and non-Accredited hospitals," *International journal of health sciences*, vol. 6, no. 7, pp. 2195–2209, Aug. 2022, doi: 10.53730/ijhs.v6ns7.11842.

[3] S. Ghosh, S. Chakraborty, N. Gupta, and S. Basu, "What ails physician review websites? A study of information needs of patients," *Decision Support Systems*, vol. 166, Nov. 2022, doi: 10.1016/j.dss.2022.113897.

[4] Y. Tian, "A review on factors related to patient comfort experience in hospitals," *Journal of Health, Population and Nutrition*, vol. 42, no. 1, Nov. 2023, doi: 10.1186/s41043-023-00465-4.

[5] B. Davis, B. K. B. -Kendall, and R. P. Dumas, "A review of COVID-19's impact on modern medical systems from a health organization management perspective," *Health and Technology*, vol. 12, no. 4, pp. 815–824, 2022, doi: 10.1007/s12553-022-00660-z.

[6] A. Das *et al.*, "Implementation of infection prevention and control practices in an upcoming COVID-19 hospital in India: an opportunity not missed," *PLoS ONE*, vol. 17, no. 5, pp. e0268071–e0268071, 2022, doi: 10.1371/journal.pone.0268071.

[7] G. Pujolar, A. O. -Anglès, I. Vargas, and M. L. Vázquez, "Changes in access to health services during the COVID-19 pandemic: a scoping review," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, Feb. 2022, doi: 10.3390/ijerph19031749.

[8] A. M. Boylan, A. Turk, M. H. V. Velthoven, and J. Powell, "Online patient feedback as a measure of quality in primary care: A multimethod study using correlation and qualitative analysis," *BMJ Open*, vol. 10, no. 2, pp. e031820–e031820, Feb. 2020, doi: 10.1136/bmjopen-2019-031820.

[9] R. Dorsey, D. Claudio, M. A. Velázquez, and P. Petersen, "Identifying service quality gaps between patients and providers in a native American outpatient clinic," *BMC Health Services Research*, vol. 22, no. 1, Feb. 2022, doi: 10.1186/s12913-022-07538-w.

[10] F. Jiang, Y. Liu, J. Hu, and X. Chen, "Understanding health empowerment from the perspective of information processing: questionnaire study," *Journal of Medical Internet Research*, vol. 24, no. 1, pp. e27178–e27178, Jan. 2022, doi: 10.2196/27178.

[11] S. Almaazmi, M. C. E. Simsekler, A. Henschel, A. Qazi, D. Marbouh, and R. A. M. A. Luqman, "Evaluating drivers of the patient experience triangle: stress, anxiety, and frustration," *International Journal of Environmental Research and Public Health*, vol. 20, no. 7, Apr. 2023, doi: 10.3390/ijerph20075384.

[12] E. Wong, F. Mavondo, and J. Fisher, "Patient feedback to improve quality of patient-centred care in public hospitals: a systematic review of the evidence," *BMC Health Services Research*, vol. 20, no. 1, Jun. 2020, doi: 10.1186/s12913-020-05383-3.

[13] F. Isasi *et al.*, "Patients, families, and communities COVID-19 impact assessment: lessons learned and compelling needs," *NAM Perspectives*, vol. 11, Nov. 2021, doi: 10.31478/202111c.

[14] R. De', N. Pandey, and A. Pal, "Impact of digital surge during Covid-19 pandemic: a viewpoint on research and practice," *International Journal of Information Management*, vol. 55, Dec. 2020, doi: 10.1016/j.ijinfomgt.2020.102171.
[15] B. J. Awrahman, C. A. Fatah, and M. Y. Hamaamin, "A review of the role and challenges of big data in healthcare informatics and analytics," *Computational intelligence and neuroscience*, vol. 2022, Sep. 2022, doi: 10.1155/2022/5317760.
[16] P. Goyal and R. Malviya, "Challenges and opportunities of big data analytics in healthcare," *Health Care Science*, vol. 2, no. 5, Oct. 2023, doi: 10.1002/hcs2.66.
[17] C. Lotfi, S. Srinivasan, M. Ertz, and I. Latrous, "Web scraping techniques and applications: a literature review," *SCRS Conference Proceedings on Intelligent Systems*, pp. 381–394, 2021, doi: 10.52458/978-93-91842-08-6-38.
[18] Ł. Augustyniak, T. Kajdanowicz, and P. Kazienko, "Comprehensive analysis of aspect term extraction methods using various text embeddings," *Computer Speech & Language*, vol. 69, 2021, doi: 10.1016/j.csl.2021.101217.
[19] H. Huang, B. Zhang, L. Jing, X. Fu, X. Chen, and J. Shi, "Logic tensor network with massive learned knowledge for aspect-based sentiment analysis," *Knowledge-Based Systems*, vol. 257, 2022, doi: 10.1016/j.knosys.2022.109943.
[20] M. V. Gopalachari, S. Gupta, S. Rakesh, D. Jayaram, and P. V. Rao, "Aspect-based sentiment analysis on multi-domain reviews through word embedding," *Journal of Intelligent Systems*, vol. 32, no. 1, 2023, doi: 10.1515/jisys-2023-0001.
[21] N. Liu and B. Shen, "Aspect term extraction via information-augmented neural network," *Complex & Intelligent Systems*, vol. 9, no. 1, pp. 537–563, 2022, doi: 10.1007/s40747-022-00818-2.
[22] H. Chouikhi, M. Alsuhaibani, and F. Jarray, "BERT-based joint model for aspect term extraction and aspect polarity detection in arabic text," *Electronics*, vol. 12, no. 3, 2023, doi: 10.3390/electronics12030515.
[23] Y. Xiong, M. M. Nor, Y. Li, H. Guo, and L. Dai, "Reinforcement learning-based aspect term extraction using dilated convolutions and differential equation initialization," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, 2023, doi: 10.14569/ijacsa.2023.0140518.
[24] A. Bansal and N. Kumar, "Aspect-based sentiment analysis using attribute extraction of hospital reviews," *New generation computing*, vol. 40, no. 4, pp. 941–960, 2022, doi: 10.1007/s00354-021-00141-3.
[25] M. Martinc, B. Škrlj, and S. Pollak, "TNT-KID: Transformer-based neural tagger for keyword identification," *Natural Language Engineering*, vol. 28, no. 4, pp. 1–40, 2021, doi: 10.1017/S1351324921000127.
[26] T. T. Goh, N. A. A. Jamaludin, H. Mohamed, M. N. Ismail, and H. S. Chua, "A comparative study on part-of-speech taggers' performance on examination questions classification according to bloom's taxonomy," *Journal of Physics: Conference Series*, vol. 2224, no. 1, Apr. 2022, doi: 10.1088/1742-6596/2224/1/012001.
[27] C. Y. Sy, L. L. Maceda, T. D. Palaoag and M. B. Abisado, "Transforming education policy: evaluating UAQTE Program implementation through LDA, BoW and TF-IDF techniques," *2024 26th International Conference on Advanced Communications Technology (ICACT)*, Pyeong Chang, Korea, Republic of, 2024, pp. 1-7, doi: 10.23919/ICACT60172.2024.10471958.
[28] E. D. Nardo, F. Polito, and E. Scalas, "A fractional generalization of the dirichlet distribution and related distributions," *Fractional Calculus and Applied Analysis*, vol. 24, no. 1, pp. 112–136, 2021, doi: 10.1515/fca-2021-0006.
[29] X. Yang and J. S. Liu, "On gibbs sampling for structured bayesian models discussion of paper by Zanella and Roberts," *Arxiv-Statistics*, pp. 1-18, 2021, doi: 10.48550/arXiv.2112.08641.
[30] A. Ganguly and S. W. F. Earp, "An introduction to variational inference," *Arxiv-Computer Science*, pp. 1-13, 2021, doi: 10.48550/arXiv.2108.13083.
[31] S. Syed and M. Spruit, "Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation," *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2017, doi: 10.1109/dsaa.2017.61.
[32] A. F. Pathan and C. Prakash, "Unsupervised aspect extraction algorithm for opinion mining using topic modeling," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 492–499, 2021, doi: 10.1016/j.gltp.2021.08.005.

BIOGRAPHIES OF AUTHORS

Nasreen Taj Mohammed Basha is working as Assistant Professor in Department of Information Science and Engineering, GMIT, Davangere. She has 18 years of teaching experience. She has completed B.E. from VTU in 2005 and M.Tech. from VTU in 2009. She has done 10 international journal publications and has attended 2 national level conferences and 8 international level conferences. Currently she is pursuing Ph.D. in computer science at Dayananda Sagar University, Kudulu Gate, Bengaluru. Her research area of interest is sentiment analysis, data analytics, and data mining. She can be contacted at email: nasreent@gmit.ac.in.

Girisha Gowdra Shivappa is working as Professor in the Department of Computer Science & Engineering, School of Engineering, DSU, Kudulu Gate, Bengaluru and has done his Ph.D. from VTU, Belagavi in the domain of Multimedia Data Mining. He has total of 24 years of academic's experience. He has 25 publications in his credit including eleven international peer reviewed journals and eight international conference papers. He has given invited lectures on advanced microprocessor, and Block chain Technology in various engineering colleges. He was involved in syllabus framing for the AMIETE courses. He is currently guiding five Ph.D. students. He has worked as coordinator in various departmental and institute level works including mentoring, organizing conferences/workshops/FDP/ technical talks, ISO, NBA & NBA accreditation. His areas of research interest include data mining, data science, and block chain. He can be contacted at email: girisha-cse@dsu.edu.in.