Why Do People (Not) Like Me?: Mining Opinion Influencing Factors from Reviews

Eda Bilici[a], Yücel Saygın[b]

[a] Institute of Digital Healthcare, WMG, University of Warwick, Coventry CV4 7AL, UK
[b] Faculty of Engineering and Natural Sciences, Computer Science and Engineering Department, Sabancı University, Orhanlı-Tuzla, Istanbul 34956, Turkey

Abstract

Feedback, without doubt, is a very important mechanism for companies or political parties to re-evaluate and improve their processes or policies. In this paper, we propose opinion influencing factors (also called as factorial aspects) as a means to provide feedback about what influences the opinions of people. We also describe a methodology to mine opinion influencing factors from textual documents with the intention to bring a new perspective to the existing recommendation systems by concentrating on service providers (or policy makers) rather than customers. This new perspective enables one to discover the reasons why people like or do not like something by learning relationships among the traits/products via semantic rules and the factors that lead to change on the opinions such as from positive to negative. As a case study we target the healthcare domain, and experiment with the patients' reviews on doctors. Experimental results show the gist of thousands of comments on particular factorial aspects associated with semantic rules in an effective way.

Keywords: Text Mining, Opinion Mining, Causality Analysis, Feedback-based Recommendations

Email addresses: `e.bilici@warwick.ac.uk` (Eda Bilici), `ysaygin@sabanciuniv.edu` (Yücel Saygın)

1. Introduction

In a decision-making process, people behave towards their aims, expectations, experiences and social interactions. Seeking causes, reasons, and explanations for various states is an important part of human nature. Nowadays, no doubt, social media become an integral part of our life and online reviews are considered as one of the richest data sources for data mining community to discover the opinion of people about various issues. However, the current focus of opinion mining community is to discover what people like or do not like about something, while in this work we intend to move opinion mining one step further by concentrating on the discovery of why people like or do not like something. For that purpose, we propose *opinion influencing factors* as a mechanism to provide feedback about what influences the opinion of people. We also propose a methodology to mine opinion influencing factors from textual documents with many possible applications. Among those applications, we have chosen recommendation systems since opinion influencing factors bring a new perspective to the existing recommender systems by providing feedback to service providers instead of customers. This is important especially for the healthcare industry since patients are increasingly using social media to write reviews and consult reviews of others about hospitals and doctors. Therefore, we have chosen healthcare as a case study and implemented our methodology on patients' reviews for doctors.

This paper presents a new methodology that aims at discovering semantic rules and the factors which cause changes in the expressed opinions. The concept of *opinion influencing factors* (also called as *factorial aspects* in the document) is introduced as a collection of aspects that have significant influence on decisions, where "aspects" are represented as collections of keywords. Learned aspects are represented as nodes in a Directed Acyclic Graph (DAG), where the directed edges represent relations between aspects that are induced from observed co-occurrence counts. Learning of aspects is based on Gibbs Sampling (also known as alternating conditional sampling) technique for La-

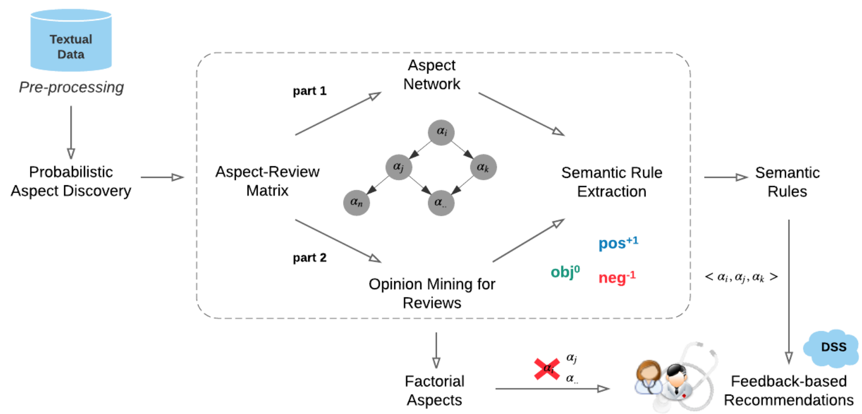

Figure 1: Overall framework of the system architecture

tent Dirichlet Allocation (LDA) which is a topic selection method. The DAG is inferred by first estimating the undirected network (*i.e.*, the moral graph) and then using a max-min greedy hill climbing search to orient the edges, based on chi-square conditional independence tests as building blocks. A bootstrap resampling strategy is used to make sure that the network structure is robust against small sampling fluctuations. Finally, semantic rules are extracted, which together with the factorial aspects are used to explain why people like or don't like something.

In Figure 1, we introduce our framework which includes six steps: (i) Data is pre-processed and aspects' keywords are extracted, (ii) Aspect network in the form of a Bayesian Network (BN) is established to obtain a graphical model, and opinion mining (sentiment analysis) is applied for each review to calculate aspect-based polarities, (iii) Semantic rules are extracted using the aspect network, and polarity degrees of them are calculated, (iv) Ordered Logit Regression technique is applied to investigate the impacts of aspects upon the opinions (*e.g.*, positive $\rightarrow$ negative), therefore, factorial aspects are determined, (v) Factorial aspects are combined with semantic rules, and finally (vi) Feedback-based recommendations are established that can be proposed by the DSS including semantic rules, and factors having significant impacts upon the opinions of people. In our study, opinion mining is used to understand the preferences of people to better serve them and to help service providers to improve themselves. Thus, service providers may have knowledge about which aspects are covered in reviews and know the reasons why the opinions of their customers change, and to which extent aspects reflect their preferences. For our experiments, we consider 406 medical doctor profiles and about 2,000 reviews retrieved from a website that doctor and hospital reviews commented by patients.

The remainder of the paper is organized as follows: In Section 2, we briefly present the related work. Then, we introduce the problem definition and preliminaries in Section 3, and probabilistic aspect discovery technique in Section 4. In Section 5, we describe the methodology. In Section 6, we introduce our novel feedback-based recommendation approach including semantic rule extraction and factorial aspect analysis. In Section 7, we discuss our experimental results, and lastly in Section 8, we conclude our study and give directions for the future research.

2. Related work

In this study, a new recommendation type called *feedback-based recommendation* is introduced including topics (aspects) that have influences upon the opinions of people, and semantic rules that are retrieved from a type of BN. Here, the related literature on belief networks and on sentiment analysis applications in healthcare are discussed. Afterwards, some related works on health recommender systems that are the part of recommender systems being applied in the healthcare industry are presented.

Networks can be designed for many purposes under varied domains such as transportation, social interaction, spreading of news, diseases, and many others. These network structures can be defined through graphs. Bayesian network (BN) also known as belief network (Zhang & Poole, 1996) is widely used as a method for the abovementioned domains that is effective on the diagnosis, prediction, classification and decision making phases. To illustrate, the natural language processing (Chapman et al., 2001), genetic diagnosis of diseases (Su et al., 2013), cancer (Zhao & Weng, 2011) and antipattern (Settas et al., 2012) detections, reliability analysis (Mahadevan et al., 2001), time-series studies (Kim et al., 2013) are some of the studies in which the BN technique is used. In this work, we introduce a novel BN application area and network type called as the "Aspect Network". We analyze patients' reviews using this network which is a graphical model that encodes probabilistic relationships among a set of aspects. Here, nodes precisely denotes aspects, and edges denote some sort of logical or discerned relationship between them.

Sentiment analysis is a trending research area which is a commonly used technique of research and social media analysis that considers extracting opinions from texts and classifying them as positive, negative or objective (Pang & Lee, 2008). Authors in (Dehkharghani et al., 2014) analyze Twitter data and apply sentiment analysis to determine the polarity degrees of texts. They establish the causality rules among aspects using a constraint-based Local Causal Discovery (LCD) algorithm. In this study, only one connection type which is the *common effect* is considered. As a part of our study, we extract rules from texts as well. Initially, we establish a type of BN called aspect network and use Max-Min Hill Climbing (MMHC) hybrid algorithm to establish the DAG. This algorithm combines constraint and score based techniques that provides more information extraction than only a constraint-based technique application. Thus, we consider three DAG connection types as *chain*, *common effect* and *common cause*. We also extract topics using a topic model like LDA but in (Dehkharghani et al., 2014), any topic extraction technique is used. Word groups are just established using the semantic distances, and topics are created without the automation. Yet, our major difference from this study is the consideration of factorial aspects, and we rely our study on their impacts upon the opinions of people associated with semantic rules. Significance, relation, cause and effect analysis between topics and opinions is a significant research area that deserve researchers' attentions. For instance, in (Li et al., 2012), a social opinion impact on topics is analyzed. Apart from them, we analyze the influence of topics on opinions. In addition, we analyze how presence/absence of one topic affects the opinions of people in reviews.

In the literature, two main topic models which are LDA (Lu et al., 2011) and Probabilistic Latent Semantic Analysis (PLSA)(Hofmann, 1999) that consider co-occurrence of words in texts, are widely studied. (Paul & Dredze, 2015) introduce SPRITE which is a set of topic models that use structured priors to create topic structures based on the users' preferences, and compare performances of several topic structures. We determine our aspects using Gibbs Sampling for LDA. This technique relies on sampling from conditional distributions of the features of the posterior. Each topic is constituted by its highest most frequent words. We choose the healthcare industry as our data source since the interest for health related issues are rapidly increasing on online platforms. In (Paul et al., 2013), patient contentment using online physician reviews is investigated, and a modified version of factorial LDA is applied to extract topics along with a sentiment analysis. In addition to this study, we include factorial aspect analysis combined with semantic rules.

Recommendation systems are designed around people's interests, needs and preferences. Content-based, collaborative filtering, demographic, knowledge-based, community-based and hybrid recommendation systems are some of the methods to find a solution for recommendation problems. (Villanueva et al., 2016) discuss semantic recommendation models and present a new semantic recommendation model called SMORE for the social media analysis. Many published studies propose healthcare-oriented recommendations. For instance, in (Zhang et al., 2013), a content-based personalized recommendation system called SocConnect is proposed, and a collaboration-based medical knowledge recommendation system for clinicians is introduced by (Huang et al., 2012). For further information on recommendation systems in healthcare, *see*, (Sanchez-Bocanegra et al., 2015; Wiesner & Pfeifer, 2014).

Users, in general, give ratings, say, from 1 to 5 under specific general titles. When service providers would like to obtain an idea about what their customers think about them, they have to read all the reviews written by their customers to have an idea if they are enough lucky. Since general titles cannot convey the whole opinions of customers, people tend to include their comments along with ratings. We extract aspects from reviews, therefore, they directly reflect real opinions of customers. None of the previous studies consider users' preferences and analyze the factors affecting their opinions as we study. As far as we are concerned, we are the first that combine semantic rules and factorial aspects for feedback-based recommendations.

3. Preliminaries and Problem Definition

To provide more insights into our methodology, we define key concepts used in this study as follows:

An "*aspect*" is associated with a group of keywords that has been commented on in reviews and "*aspect lexicon*" is a set of aspects with associated keyword list for each aspect for a given domain. Here, we introduce a new concept "*opinion influencing factors*" also called as "*factorial aspects*" that refers to the significant aspects, in other words, aspects having impacts upon the opinions of people. When an aspect and its sentiment (opinion) appear in one review, we call them as "*aspect-sentiment pair*". A "*sentiment value*" is a score that takes values between -1 and 1, measuring the polarity of a sentiment. Sentiment values can be categorized as positive, negative and neutral (objective) where 1 denotes the most positive sentiment, -1 denotes the most negative one and the polarity of neutral (objective) sentiment can be around 0. The following statement would be a nice instance to define a positive tagged sentence: "*Dr. X is a very knowledgeable doctor I will go again*". Here, "Knowledge" refers to an aspect. "Knowledgeable" refers to the sentiment bearing aspect, and "very knowledgeable doctor" refers to its sentiment representing an aspect-sentiment pair that defines a positive sentiment on the knowledge. In this paper, *aspect-based sentiment analysis* is performed with the lexicon technique. Thus, we create our lexicon using LDA and WordNet (Miller, 1995), then per-form aspect-based sentiment analysis for texts. In our domain, an opinion is a subjective statement describing what a patient thinks about a doctor and/or service. We calculate polarity scores using AlchemyAPI sentiment analysis tool (*see*, www.alchemyapi.com) for each review. These scores are then converted to tags and associated with corresponding semantic rules.

Definition 1. Let $\{\alpha_1, \alpha_2, ..., \alpha_n\}$ be the set of n aspects, $i = 1, 2..., n$. Each aspect has its own keyword group, and a keyword of aspect i does not appear in any other aspects. $\{\theta_{11}, \theta_{12}, ..., \theta_{nv}\}$ be the set of v keyword groups of n aspects. $\{\omega_{111}, \omega_{112}, ..., \omega_{nvt}\}$ be the set of t keywords, and ω_{ihq} denotes the q^{th} keyword in the keyword group h $(\in v)$ of the aspect i, $q = 1, 2, ..., t$. $\{r_1, r_2, ..., r_m\}$ be the set of m reviews in which each review r_y includes a set of aspects associated with a set of keyword groups and a set of keywords, $y = 1, 2, ..., m$.

Semantic stands for the meaning of phrases and words. We use this concept and frequent word patterns to group the keywords, and each keyword group is associated with its related aspect. Using this information, aspect network which is a kind of BN, presents an interaction between probability and graph theory including a set of conditional independence relationships summarized through graphs is established. In our study, the gist of reviews are represented by aspects that are shown in the form of graphs.

Definition 2. Let $G = \{V, E\}$ be the directed acyclic graph (DAG) where V and E stand for the set of vertices (nodes) also called as aspects where $\{\alpha_1, \alpha_2, ..., \alpha_n\}$, and edges (arcs) that refer to the set of ordered pairs of vertices, respectively. *Dependence(d)-separation* is a measure to determine from a given DAG if an aspect α_i is independent of another aspect α_j given a third aspect α_k. If α_i and α_j are connected by an edge, then α_i and α_j are dependent. In other words, whether G is a DAG where two aspects α_i and α_j are *d-separated* given a third aspect α_k in G, then they are conditionally independent on α_k.

All paths between α_i and α_j are d-separated by α_k that can be represented as $\alpha_i \perp\!\!\!\perp \alpha_j \,|\, \alpha_k$. α_i and α_j are conditionally dependent given α_k *iff* information about one aspect affects the opinions about the other under α_k. Likewise, α_i

and α_j are conditionally independent given α_k *iff* information about one aspect does not affect the opinions about the other under α_k, $i, j, k = 1, 2, ..., n$.

Definition 3. Let $\{\gamma_1, \gamma_2, ..., \gamma_f\}$ be the set of f semantic rules, where γ_p refers to a semantic rule that includes triple aspect dependencies (or also called as directed paths) $< \alpha_i, \alpha_j, \alpha_k >$, $p = 1, 2, ..., f$, and triple aspect dependencies can be in the form of four directed paths based on d-separations in a DAG as follows: (i) $\alpha_i \rightarrow \alpha_j \rightarrow \alpha_k$ be a directed path from α_i to α_k through α_j where α_i is an indirect cause of α_k, and $\alpha_i \leftarrow \alpha_j \leftarrow \alpha_k$ be a directed path from α_k to α_i through α_j where α_k is an indirect cause of α_i. These connection types stand for *chain* connections. In both cases, α_i and α_k are conditionally independent given α_j, (ii) $\alpha_i \leftarrow \alpha_j \rightarrow \alpha_k$ be a pair of directed paths from α_j to α_i and α_j to α_k where α_j is a *common cause* of α_i and α_k. These abovementioned paths have causal relations that brings about dependence between α_i and α_k, and lastly, (iii) $\alpha_i \rightarrow \alpha_j \leftarrow \alpha_k$ be a directed path where α_i and α_k have a *common effect* in α_j, yet there is no causal relation between them.

Aspect triples are determined based on the co-occurrences of aspects in reviews. Information about the dependence relationships of aspects are employed to extract rules. In our study, not all the aspects have significant impacts upon the opinions of people. For this reason, we extract the aspects that the occurrence of them in reviews change the polarity of the reviews. In our context, these aspects are defined as opinion influencing factors and called as *factorial aspects* (FAs). When these aspects occur in reviews, the opinions of people change say from positive to negative.

Definition 4. Let α_i be the factorial aspect that has an effect on opinions where $\alpha_i \in \{\alpha_1, \alpha_2, ..., \alpha_n\}$. Because our dependent variable (*i.e.*, polarity of each aspect or review) is ordinal and have three categories, Ordered Logit Regression statistical technique is used to determine the FAs that measure the relationship between a dependent variable (outcome tag) and independent variables (aspects) by predicting probabilities using a logit link function.

To summarize, a review r_y includes a set of n aspects associated with a set of s keyword groups. Each keyword group of aspect i includes a set of t words. First, aspect network is established without any information regarding the impacts of aspects upon the opinions. This network is formed by the co-occurrences of aspects in reviews. Opinion mining is applied to determine the polarity degrees of each aspect i in the set of n aspects. Polarities are assigned to each aspect i. Semantic rules are established, and then polarity degrees for each rule are assigned as well. Here, we have no information on whether or not a single aspect has an impact upon the opinions of people. For this reason, FAs and their contributions on opinions are determined using the Ordered Logit Regression analysis. This information is used as an input to select appropriate semantic rules, *i.e.*, $< \alpha_i, \alpha_j, \alpha_k >$. Finally, feedback-based recommendations are proposed that include the joint analysis of factorial aspects and semantic rules.

4. Probabilistic Aspect Discovery

Initially, we apply the pre-processing step to clean and prepare the data for the analysis. We have finally 1,832 patients' reviews. After we determine the frequency of keywords occurred (*e.g.*, top 10 words) per aspect, we decide the suitable number of clusters using Gibbs Sampling technique which is an algorithm from the family of Markov Chain Monte Carlo (MCMC) framework. In this section, the data preparation, keyword extraction, and aspect selection method which is Gibbs Sampling for Latent Dirichlet Allocation are discussed.

4.1. Pre-processing

The vocabulary may include many unrelated words which do not contribute the considered aspect structure of the corpus and may deteriorate the models' ability to find topics. In order to select proper vocabularies, pre-processing is required such as stemming the words, and removing stopwords, punctuations, numbers to increase the predictive power of the study. After the pre-processing, we have 1,832 reviews with 665 words. We use R text mining package "tm" (*see*, http://tm.r-forge.r-project.org) for this pre-processing stage. Afterwords, we transform the dataset into a document-term matrix for the LDA analysis.

4.2. Learning aspects with Gibbs sampling

Latent Dirichlet Allocation (LDA) is a widely used probabilistic topic model in which each document is modeled as a mixture over the latent topics, and each topic has a multinomial distribution over the entire vocabulary, in other words, a collection of data namely corpus (Blei et al., 2003). We use R package "topicmodel" (Grün & Hornik, 2011) that provides Gibbs sampling technique for LDA. In this study, Gibbs sampling is used as a standard estimation method. We generate several topics, and each topic includes several words ordered by the number of times that word assigned to the topic. Words are associated with the selected topics, and grouped using semantic distances (*i.e.*, degree of similarity of words) between synsets in WordNet (Budanitsky & Hirst, 2006), which is a lexical database like a thesaurus (*see*, https://wordnet.princeton.edu). Each topic includes a bag of words, and these topics are called as *aspects* in our domain. Common words in topics are removed since each topics' keywords should be unique. In other words, each topic is independent from the other topic and includes unique word groups. Finally, 10 topics are chosen, and keyword lists are constituted.

5. Methodology

In this section, aspect network, learning in the aspect network, measures of aspect connections, and aspect-rule tag classifications are discussed, respectively. Analyzing reviews and comments in terms of their graphical structures enable substantial insights. When we view the reviews as a graph, it provides us a better understanding of the logical relationships in reviews defined by nodes with its associated links.

Table 1: Aspect-review matrix including 1,832 reviews covering 10 aspects

#	Helpfulness	Concern	Diagnosis	$\cdots$	Staff
1	1	1	0	$\cdots$	1
2	0	0	1	$\cdots$	0
3	1	0	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
1,832	1	1	0	$\cdots$	1

5.1. Aspect network.

Aspect network is a type of Bayesian network which is a directed acyclic graph (DAG), $G = \{V, E\}$ that consists of a set of n vertices (nodes) in $V = \{\alpha_1, \alpha_2, ..., \alpha_n\}$, and in our context, we call vertices as aspects, and a set of edges (arcs) in E that denotes the conditional independence relationships between some pairs of aspects using the presence or absence of direct causations, for further information on BNs, *see*, (Pearl, 2000). The joint probability distribution of the set of n aspects in the aspect network can be defined as:

$$P(\alpha_1, \alpha_2, ..., \alpha_{n-1}, \alpha_n) = \prod_{i=1}^{n} P(\alpha_i \mid Pa(\alpha_i)) \tag{1}$$

where $Pa(\alpha_i)$ denotes the set of parent nodes of the aspect i in G. To explain and illustrate our method, we introduce six aspects extracted from patients' reviews and these are Helpfulness (H), Kindness (K), Listener (L), Diagnosis (D), Knowledge (W) and Concern (C), *see*, Figure 2. Reviews are converted into the aspect-review matrix, and the aspect set of 6 aspects $\{\alpha_1, \alpha_2, ..., \alpha_6\}$, where the components α_i are either 0 or 1 denoting the absence/presence of the corresponding aspect in the aspect network, $i = 1, 2, ..., 6$.

Aspect-review matrix. After aspects are extracted with their corresponding keyword groups and words, we are able to create an aspect-review matrix as Table 1. Formally, we define the matrix as a set of n aspects $\{\alpha_2, \alpha_2, ..., \alpha_n\}$ and each aspect is associated with its keyword groups. Let $\{\theta_{11}, \theta_{12}, ..., \theta_{nv}\}$ be the set of v keyword groups of n aspects where θ_{ih} denotes the keyword group h of aspect i where $i = 1, 2, ..., n,\ h = 1, 2, ..., v$. Each review in the set of m reviews $\{r_1, r_2, ..., r_m\}$ includes the set of e ($\in n$) aspects, and each aspect in the review r_y is either 1 (*i.e.*, if any keyword in its corresponding keyword group of aspect i appears in review r_y) or 0 (*i.e.*, if any keyword does not appear in its corresponding keyword group of aspect i in review r_y). For instance, while two aspects can be appeared in review x, four aspects can be appeared in review y as follows: $r_x = \{\alpha_1, \alpha_2\}$ and $r_y = \{\alpha_1, \alpha_2, \alpha_5, \alpha_6\}$, respectively where $x, y, = 1, 2, ..., 1, 832$.

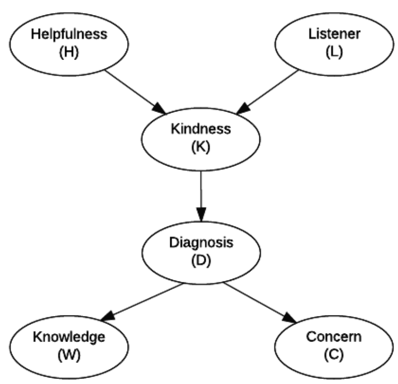

Figure 2: A simple aspect network representing connections among six aspects

Separations in a graph refer independence relations in a probability distribution, and particular independence relations can be constructed using *d-separations* in the related DAG.

Causal graphs. Graphical connections in DAGs can be shown through three different types of triples (Salmon, 1980): common cause, chain, and common effect. If aspect K is the cause of both aspect H and aspect L, this connection refers to *common cause connection*. H and L are conditionally independent given K and the notation for independence can be shown as $H \perp\!\!\!\perp L \,|\, K$. When K is known,

K separates (or blocks) the flow between H and L. The joint density can be expressed as $P(H,K,L) = P(H\backslash K)P(L\backslash K)P(K)$ and shown as $H \leftarrow K \rightarrow L$. If the occurrence of aspect H causes K, and K causes L, this connection refers to *chain connection*. Aspects H and L are independent given the aspect K, the notation for independence can be shown as $H \perp\!\!\!\perp L \,|\, K$. K separates the flow from H to L. In other words, there is no direct flow between H to L. The joint density can be expressed as $P(H,K,L) = P(L\backslash K)P(K\backslash H)P(H)$ and can be shown as $H \rightarrow K \rightarrow L$. If one aspect has two parents which are independent except if the child is given, this connection refers to *common effect connection* (*v-structure*). Both aspects H and L are independent and they become dependent as K is known. The flow between H and L is separated (or blocked) when K is not observed. Aspects H and L are conditionally independent, and the notation for independence can be shown as $H \not\perp\!\!\!\perp L \,|\, K$, but independence depends on the information flow on K. The joint density can be expressed as $P(H,K,L) = P(K\backslash H,L)P(H)P(L)$, and can be shown as $H \rightarrow K \leftarrow L$. The network that we consider is acyclic; in other words, aspect relations cannot have any loops as $H \rightarrow K \rightarrow \cdots \rightarrow H$ or bi-directional as $H \leftrightarrow K$. In this study, we analyze triple aspect relations. Let's say, we investigate the probability of commenting on two aspects H and L together, and what is the probability of commenting on aspect K as well? H and L are conditionally independent given K and the notation for independence can be shown as $H \perp\!\!\!\perp L \,|\, K$. Patients comment on doctors via online social platforms, we would like to know, for example, what are the reasons of patients to comment on a doctor(s)? Here, reasons denote our aspects in which we establish them using Gibbs sampling for LDA topic selection technique, and each aspect has a keyword group behind. We use Bayes' theorem to calculate the posterior probabilities of the aspects. Figure 2 shows a partial aspect network representation of patients' reviews. The joint density of these six aspects can be defined as:

$$P(H,L,K,D,W,C) = P(K\backslash H,L)\,P(D\backslash K)\,P(W\backslash D)\,P(C\backslash D)\,P(H)\,P(L) \quad (2)$$

For instance, we're interested in *Kindness* aspect, and would like to analyze the probability of associations with other aspects, say, *Helpfulness*. We refer to $P(H)$ as the prior probability of *Helpfulness* because it expresses our understanding of the probability of H without any information about whether *Kindness* has occurred. Similarly, we define $P(K\backslash H)$ as the posterior probability of H given K because it expresses our understanding of the probability of H that we know that K has occurred. The effect of knowing K is, therefore, defined in the change from the prior probability of H to the posterior probability of H.

5.2. Learning

Learning in the aspect network has two main steps: (i) learning the structure of the network, and (ii) learning the parameters. Establishing the graphical structure which presents the conditional independencies refers to the *structure learning* whereas in the *parameter learning* phase, parameters of the local distribution are estimated using the framework obtained in the learning phase.

In the literature, three main applications have been developed to learn the structure of Bayesian networks from data; constraint-based, score-based and hybrid algorithms. To provide more insight into our application, we briefly discuss these three methods used in the literature: (i) *Constraint-based algorithms* (Schlüter, 2014) learn the undirected graph (skeleton) of an underlying Bayesian network using conditional independence tests to discover the Markov blankets (dependencies) of the nodes. The rejection of the conditional independence determines the related d-separation that should be exist in the network. *The Local Causal Discovery (LCD)* algorithm (Mani & Cooper, 2004) is one of the widely applied constraint-based method. *The Grow-Shrink (GS)* (Margaritis & Thrun, 2000), *the PC* (Li & Shi, 2007), *the Fast Causal Inference (FCI)* (Colombo et al.,

procedure: Max-Min Hill Climbing (**Aspect-matrix**)
Input: **Aspect-matrix**
Output: DAG on the aspects in **Aspect-matrix**
%Restrict
for every aspect $\alpha_i \in \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ **do**
| $\boldsymbol{PC}_{\alpha_i}$ = Max-Min Parent Child (α_i , **Aspect-matrix**)
end
%Search
Initiating from an empty graph apply greedy hill-climbing
with edge orientations. *add-edge* $\alpha_j \rightarrow \alpha_i$ if $\alpha_j \in \boldsymbol{PC}_{\alpha_i}$
return the highest score DAG achieved
end procedure

Figure 3: The Max-Min Hill-Climbing algorithm of Tsamardinos et al. (Tsamardinos et al., 2006)

2012), and *the Incremental Association Markov Blanket (IAMB)* (Tsamardinos et al., 2003) are some of the other well-known constraint-based algorithms in the literature, (ii) *Search and score based algorithms* (Acid et al., 2013) search all the space and assign a score to each structure and choose the structure with the highest score. Heuristic-based approaches like *Hill-Climbing (HC)* (Gámez et al., 2011), and *Genetic Algorithm (GA)* (Larrańaga et al., 1996) are some of the well-known techniques under this category, and lastly, (iii) *Hybrid algorithms* use both constraint based and search and score based techniques to establish the graph. Initially, they use constraint-based techniques to establish the skeleton of the graph applying conditional independence tests to confine the search space, then identify the orientation with search and score based techniques.

We consider a hybrid algorithm of Tsamardinos et al. (Tsamardinos et al., 2006) which is called Max-Min Hill Climbing (MMHC) using "bnlearn" (Scutari, 2009), an open source software package in the statistical computing tool R (*see*, http://www.r-project.org) to learn the aspect network structure. In Figure 3, the steps of the algorithm is described in detail. MMHC begins with the constraint-based local causal discovery algorithm called Max-Min Parent Child (MMPC) algorithm to establish the undirected graph (skeleton) of an underlying aspect network. A greedy Bayesian-scoring hill climbing search is employed in order to orient (*e.g.*, add, delete and remove) the edges and find the optimal aspect network. Conditional independence (d-separation) tests are applied to present relations between aspects. Since we consider a hybrid algorithm, we have to compute network scores as well as conditionally independence test in the parameter learning phase. In order to learn the aspect network, we employ *Pearson's* χ^2 as a conditional independence test with 95% confidence (α= 0.05) that measures the associations and the strength among aspects. Because parameters are learned conditional on the results of structure learning, we employ *model averaging* approach combining with a nonparametric bootstrap that averages predictions over bootstrap samples to get a robust network from the data. Network structure is learned from each bootstrap sample with a Max-min Hill Climbing search, and to compute model likelihoods, *Bayesian Information Criterion (BIC)* is used as a scoring technique. Links are considered significant if they occur in at least $\geq$ 50% of the network. This is our minimum support value and below this value our output does not change. The strength of the edge and the degree of confidence of the direction of the aspect connections using non-parametric bootstrap algorithm can be computed as follows: For instance, say, aspects $\alpha_i \rightarrow \alpha_j$ occurs g_1 times and $\alpha_j \rightarrow \alpha_i$ occurs g_2 times in the $\mathcal{G}$ network, the bootstrap edge strength between α_i and α_j can be computed as $(g_1 + g_2)/\mathcal{G}$, $i, j = 1, 2, ..., n$. Combination of bootstrap models using averaging scheme to obtain an averaged model provides us a stable structure.

5.3. Aspect-rule tag classification

In this section, we introduce our tag classification steps for each aspect and rule. Initially, polarity values for each aspect and rule are calculated using the AlchemyAPI. Thus, each review has its own score. To categorize polarities of reviews, pre-determined threshold value which is $\pm$ 0.1 is chosen. Polarity assignments are also called as *tag classification* where denoted as $Tag(\mathcal{T}) = \mathcal{T}_{\mathcal{P}} - \mathcal{T}_{\mathcal{N}}$ denotes the polarity of the review, in other words, the class of opinion. $\mathcal{T} \in [-1, 1]$, {negative, objective, positive}. $\mathcal{T}$ can be defined as follows: if $\mathcal{T} \in [-1, -0.1)$, then tagged as negative, if $\mathcal{T} \in [-0.1, 0.1]$, then tagged as objective and if $\mathcal{T} \in (0.1, 1]$, then tagged as positive. In order to tag an aspect, we choose the selected aspect, say, α_i and then we tag each review that the selected aspect has occurred. Similarly, we choose a semantic rule retrieved from the aspect network, say, $< \alpha_i, \alpha_j, \alpha_k >$ that three aspects co-occur in reviews and then we tag each review that these three aspects belong to, $i, j, k = 1, 2, ..., n$. Note that we only tag a rule *iff* aspect triples in this rule include factorial aspects.

6. Feedback-based Recommendations

Feedback-based recommendations consist of two parts: aspect-based semantic rule extraction, and factorial aspect analysis. Because the aspect network has no information about the degree of opinions, we do not know whether or not the aspect appeared in reviews is significant. If an aspect is not significant, it cannot be a factor. Aspect triples can only be considered as a rule if they pass the conditional independence test, their association is greater than the minimum support level and aspects in the rule are factorial. Here, aspect share and polarity-based aspect frequency calculations are introduced to provide more understanding for our methodology. First of all, aspect frequencies are calculated for each aspect using with the following formula:

$$\omega_i = \frac{\sum_{i=1}^{n} \alpha_i}{\mathcal{R}} \tag{3}$$

where ω_i denotes the aspect frequency of aspect i in the set of m reviews. $\mathcal{R}$ be the set of all reviews where $\mathcal{R} = \{r_1, r_2, ..., r_m\}$, and α_i denotes the aspect i that has appeared in reviews, $i = 1, 2, ..., n$. To compute the polarity-based aspect share of aspect i that has appeared in positive/objective/negative tagged reviews, the following formulation is used:

$$\vartheta_i = \frac{\sum_{i=1}^{n} \alpha_i}{\mathcal{R}^{-/\circ/+}} \tag{4}$$

where ϑ_i is the polarity-based aspect shares of aspect i. $\mathcal{R}^-$, $\mathcal{R}^\circ$ and $\mathcal{R}^+$ refer to the set of negative, objective and positive tagged reviews, $\{\mathcal{R}^-, \mathcal{R}^\circ, \mathcal{R}^+\} \in \mathcal{R}$.

6.1. Semantic rule extraction

Semantic rule γ_p ($\in \mathbf{f}$) be the aspect triple $< \alpha_i, \alpha_j, \alpha_k >$ that selected based on aspect co-occurrences in reviews, and co-occurrence information is extracted using *d-separations* in the aspect network, *see* 5.1. Afterwards, polarities for each semantic rule p is assigned. The polarity percentages of each rule can be calculated using the following formula:

$$\Phi_p = \sum_{i,j,k=1}^{n} \frac{\gamma_p^{-/\circ/+}}{M_{ijk}} \tag{5}$$

where Φ_p denotes the polarity percentage of rule p, $p = 1, 2, \ldots, f$. $\gamma_p^{-/\circ/+}$ denote the number of negative, objective and positive tagged rules inferred from the combination of aspects i, j and k. M_{ijk} denotes the number of reviews that aspect i, j and k have co-occurred in the reviews, $i, j, k = 1, 2, \ldots, n$. For instance, $\alpha_i \leftarrow \alpha_k \rightarrow \alpha_j$ or $\alpha_k \rightarrow \alpha_i, \alpha_j$ is a connection type and can be considered as a rule like $< \alpha_i, \alpha_j, \alpha_k >$, *see*, Section 5.1 for more information on graphical aspect connections.

6.2. Factorial aspect analysis

Sentiment analysis of reviews is a regression problem, where there is a number of independent variables, that when taken together, produce a result namely a dependent/outcome variable. In this study, we consider 10 aspects that refer to independent variables and each of them has appeared in a review. Each aspect has its own "tag" with three ordinal opinion categories as negative (1), objective (2), and positive (3). We establish an *ordinal logistic regression model*, also called as *ordered logit model* and analyze it using Minitab 17.

Definition 5. Let $\mathcal{T}$(tag) be the outcome variable denoting the opinions with the opinion class set $\mathbf{s}$ = {negative(1), objective(2), positive(3)} that are conditional on the components of aspect set $\{\alpha_1, \alpha_2, ..., \alpha_n\}$ and the values realize with probabilities $P_1, P_2, \ldots, P_s$. $\mathbf{z}$ stands for the vector of a constant term and n aspects (covariates).

Initially, we determine which tag class to employ as the base value. Outcome of interest is conditional on a distinct value (presence or absence) of the aspect. Ordered Logit model predicts the logit of $\mathcal{T}$ from the vector $\mathbf{z}$. We have two logit link functions for the three tag classes. For instance, we choose $\mathcal{T} = 1$ (negative) be the outcome to constitute logit link functions comparing this outcome with other tag classes. Two logit link functions can be computed as follows:

$$l_c(\mathbf{z}) = ln\left\{\frac{P(\mathcal{T} = c \mid \mathbf{z})}{P(\mathcal{T} = 1 \mid \mathbf{z})}\right\} = \beta_{c0} + \beta_{c1}\alpha_1 + ... + \beta_{cn}\alpha_n \tag{6}$$

where c refers to the class of the logit link function and subset of the opinion class set $\mathbf{s}$, c=2 (objective), 3 (positive). β_{c0} be the constant term and intercept of the $\mathcal{T}$, and β_{cn} be the slope and regression coefficient and shows the direction of the relationship between aspect and the logit of opinion. In Equation 6, logit of opinions in class c are compared to negative tagged opinions conditional on each aspect in the aspect set. The conditional probabilities of each tag class s given $\mathbf{z}$ can be shown as follows:

$$P(\mathcal{T} = s \mid \mathbf{z}) = \frac{e^{l_s(\mathbf{z})}}{1 + e^{l_2(\mathbf{z})} + e^{l_3(\mathbf{z})}} \tag{7}$$

where $l_1(\mathbf{z}) = 0$. The *odds ratio* (π_{ci}) be the probability of realizing the outcome of interest explains the change in odds of $\mathcal{T}$ given a unit change in the aspect set $\{\alpha_1, \alpha_2, ..., \alpha_n\}$ where the components of the set are either 0 or 1. We choose the outcome tag as *negative* (1). So, the odds ratio of $\mathcal{T} = c$ versus outcome tag $\mathcal{T} = 1$ for aspect values of $\alpha_i = 1$ (presence) vs $\alpha_i = 0$ (absence) in reviews, where $\alpha_i \in \mathbf{z}$ can be computed as follows:

$$\pi_{ci}(1,0) = \frac{P(\mathcal{T} = c \mid \alpha_i = 1)/P(\mathcal{T} = 1 \mid \alpha_i = 1)}{P(\mathcal{T} = c \mid \alpha_i = 0)/P(\mathcal{T} = 1 \mid \alpha_i = 0)} \tag{8}$$

The aim to use the ordered logit model can be summarized as follows: (i) Determining the significant aspects that have an effect on the ordinal opinion, (ii) Analyzing the validity of the regression model and classes of opinions, and (iii) Explaining the direction of the relationship between aspects and the opinions. In this paper, we consider three classes of opinions associated with the (non) occurrence of 10 aspects in reviews. In the results and experiments section, details of the analysis are provided.

7. Experiments & Results

In this section, experiments and their results are discussed. Initially, accuracies of tag classifications are tested using several machine learning methods. Polarity degrees of each aspect are presented, and the results of logit model including aspect-sentiment pairs to determine factorial aspects and to quantify the impacts of aspects on decisions are evaluated. Then, aspect network with corresponding semantic rules is introduced, and lastly, semantic rules combined with factorial aspects along with summary statements that form the feedback-based recommendations are presented.

7.1. Results

After the application of sentiment analysis, polarities are assigned for each aspect. We have three (ternary) types of review classifications having negative, objective and positive sentiments. Accuracies of tag classifications are tested using two supervised learning algorithms as Naive Bayes (NB) which is a generative method, and Support Vector Machine (SVM) which is a robust discriminative method with 10-fold cross validation. Weka, a suite of machine learning software written in Java, developed at the University of Waikato is used for the classifications. Classification results are 69% and 67%, respectively. Accuracies of these classifiers are slightly higher than 70%, if we exclude objective tagged reviews.

As a result, we have 37% negative, 4% objective and 59% positive tagged reviews. Thus, we can deduce that people have substantially commented positively on doctors and/or their services. Our focus is especially on positive and negative commented reviews since the objective commented reviews are neutral, in other words, presence or absence of the aspect(s) have no influence on the opinions. Figure 4 indicates the aspect frequencies and aspect polarity shares in overall reviews. We refer readers to Equation 4 and Equation 5 for the aspect frequency and polarity share calculations, respectively. While the aspect *Concern* has the highest frequency (46%), the aspect *Professional* has the lowest (14%) frequency in reviews. Do you think the frequency of words in reviews are enough to reach a decision on the opinions of people? Of course, the answer is NO! But, Why?

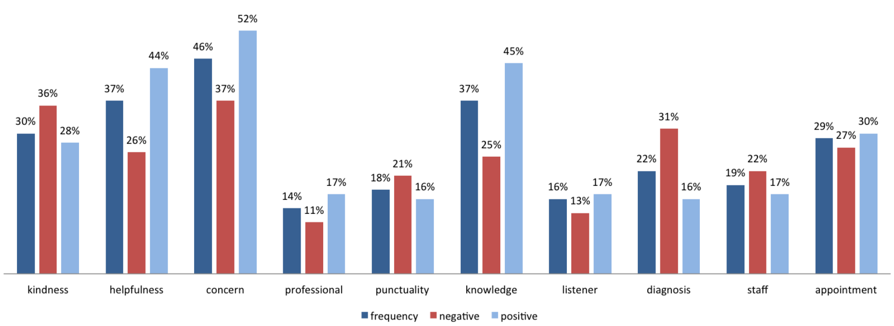

Figure 4: Aspect frequencies and polarities of overall reviews

For instance, patients are likely to say that “if the doctor is very knowledgeable, his X aspect is not important for me”. Here, X is taken into account for the frequency calculation but it has no impact on the opinions. Polarity-based aspect shares denote the polarity shares in terms of percentages in overall reviews. The impacts of aspect *Concern* and *Professional* are almost same. To

Table 2: Summary of ordered logit regression model

Predictor	Coef.	SE Coef.	Z	P	Odds ratio	95% CI Lower	95% CI Upper
Constant(1)	0.203	0.126	1.61	0.108			
Constant(2)	0.397	0.127	3.13	0.002			
Kindness	0.262	0.111	2.37	0.018	1.30	1.05	1.61
Helpfulness	-0.887	0.110	-8.09	0.000	0.41	0.33	0.51
Concern	-0.671	0.104	-6.46	0.000	0.51	0.42	0.63
Appointment	-0.280	0.115	-2.44	0.015	0.76	0.60	0.95
Professional	-0.615	0.152	-4.04	0.000	0.54	0.40	0.73
Punctuality	0.233	0.129	1.80	0.072	1.26	0.98	1.63
Knowledge	-0.898	0.109	-8.23	0.000	0.41	0.33	0.50
Listener	-0.295	0.144	-2.05	0.040	0.74	0.56	0.99
Diagnosis	0.713	0.121	5.88	0.000	2.04	1.61	2.59
Staff	0.468	0.129	3.63	0.000	1.60	1.24	2.05

analyze the impacts of aspects on the opinions, we conduct an ordered logit analysis that defined in Section 6.2. Polarities are calculated for the each aspect and rule, therefore, we can easily use this information as a an input to reach a decision on what patients like or do not like about the doctor and/or his service, and find out the reasons behind their (dis)contentment.

Summary of ordinal logit regression statistics including the estimated coefficients, standard error of the coefficients, z-values, p-values, odds ratios and 95% confidence interval for the odds ratio are presented in Table 2. Two tail p-value test the hypothesis that each coefficient is different than zero. The p-value has to be less than the threshold level ($\alpha = 0.05$) to reject the null hypothesis, and say, the aspect has a significant impact upon the opinion. Constant(1) and Constant(2) are predicted coefficients that obtained from each logit link function, *see*, Equation 6. For a given aspect with a 0.05 confidence, we would say that we are 95% confident that CI shows an interval in which the proportional odds ratio would take place. Opinions of people denote the ordinal outcome variable with three classes. Odds refer to the combined effect on the classes of opinions. Odds ratio is used to compare the effects of one unit change in the selected aspect on the classes of opinions given the other aspects are held constant in the model. Positive coefficient shows that a one unit increase (presence) (*i.e.,* $0 \rightarrow 1$) of an aspect i, and an odds ratio that is greater than 1 shows that the aspect is more likely to be associated with the first category of opinion which is negative, $i = 1, 2, ..., 10$. Similarly, negative coefficient shows that higher categories are more likely.

For instance, the coefficient (β) of 0.262 for *Kindness* is the predicted change in the logit of the cumulative opinions probability comparing a one unit change in the aspect on the classes of opinions given the other aspects are held constant in the model. Since the p-value for the predicted coefficient is 0.018, there is sufficient evidence to conclude that *Kindness* has an impact upon opinions. The proportional odds ratio for a one unit change in *Kindness* results in a 30% ($e^{0.262}$=1.30 times) increase in the odds that people have negative opinions versus the combined opinion classes as objective and positive and that the combined opinion classes as negative and objective versus positive opinions given that all of the other aspects in the model are held constant. Since the p-value for estimated coefficient of *Punctuality* is 0.072, there is insufficient evidence to conclude that this aspect has an impact upon opinions of people. The p-values for estimated coefficients of other aspects are less than the significance level, α= 0.05, and there are sufficient evidences to conclude that aspects (except *Punctuality*) influence patients' opinions. In total, we have 680 negative, 73 objective and 1,079 positive tagged reviews. Thus, we have 862,127 $((680 * 73) + (680 * 1,079) + (73 * 1,079))$ opinion pairs. Using ordered logit analysis, we find that 70.3% of pairs are concordant that also support the tag classification results of NB and SVM.

Our aspect network is learned by the Max-Min Hill Climbing hybrid al-

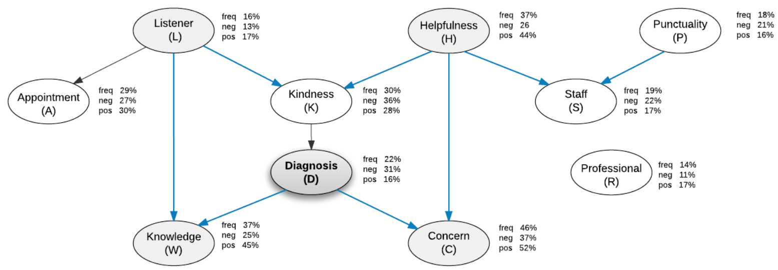

Figure 5: Aspect network of overall reviews

gorithm. Max-Min Parent Children (MMPC) is used as a constraint-based method, and Bayesian Information Criterion (BIC) is used to compute model likelihoods. Pearson's χ^2 is used as a conditional independence test. The alpha threshold is chosen as 0.05. We use R package "Rgraphviz" for graphical representations of the aspect network. We refer readers to Section 5.1 for further information on interpretation of the aspect network. We repeat the structure learning phase several times with different initializations to decrease the effect of having the locally optimal networks. Afterwards, we average the learned structure to obtain a more stable network. We predict the confidence threshold for all possible edges for 100 nonparametric samples and this minimum support threshold is determined as $\geq 50\%$ that denotes the strength of each edge, can be accepted as a significance value for the averaged network. The confidence in the direction of the edges is calculated as the probability of the certain direction in the bootstrap replications given the existence of an edge between from one aspect to another one. Aspect network is presented in Figure 5 where blue arrows denote the *v-structures*. It is explicit that only the aspect *Professional* has no relations with other aspects.

7.2. Feedback-based recommendations

To establish recommendations for service providers, we use two main information that retrieved from the aspect network and factorial aspect analysis. Ordered logit regression is used as a factor analysis method enabling us to know the significant aspects upon the opinions of people. Hence, we can exclude insignificant ones from our model. In our case, only the aspect *Punctuality* has no significant impact upon the opinions, therefore, we exclude it from the further analysis. Odds ratio in factor analysis shows the impact of one unit change in an aspect that is independent of the values of the other aspects. We now have the information on the directions and the magnitudes of the relationship between the aspects and the classes of opinions.

In our study, our focus is on aspects that their occurrence in reviews have higher impacts on negative opinions more than positive ones. Thus, service provider can easily better his service using this information. We choose the aspect *Diagnosis* that occurrences in reviews has the highest negative impact on the opinions of patients (*e.g.*, positive $\rightarrow$ negative). A one unit change in *Diagnosis* results in a 2.04 times increase in the odds that an opinion is negative versus the combined objective and positive classes of opinions and that the combined negative and objective versus positive level of opinions given all other aspects are held constant. The impact of the *Diagnosis* in reviews are obvious and the occurrence of this aspect has higher influence on negative opinions than positive ones. For instance, *Helpfulness*, *Knowledge*, *Concern* and *Listener* aspects are statistically significant in our logit analysis, and they highly exist in positively tagged reviews. Yet, their triple relations show different polarity degrees. As we have discussed before, we use the ordered logit regression analysis to determine the significant factors, to validate the model and interpret the magnitudes and relationships of the directions between aspects and the classes of opinions, and then we use this information as an input to establish semantic rules.

In Table 3, selected rules along with rule polarities are shown. The first three columns indicate the aspect relations and their types of connections. The last two columns indicate the highest polarity degree of the rule and its related tag. How can we interpret the extracted semantic rules? When we consider semantic rules with their associated polarities, we can easily see that aspects

Table 3: Selected rules extracted from the aspect network

#	Rules	Aspect Triples	Con. Type	Polarity%	Tag
1	**D**, H → C	**D**, C, H	*com. effect*	66	pos
2	L, **D** → W	**D**, W, L	*com. effect*	64	**neg**
3	**D** → W, C	**D**, W, C	*com. cause*	67	pos
4	H → K → **D**	**D**, K, H	*chain*	50	**neg**
5	L → K → **D**	**D**, K, L	*chain*	54	pos

and their relations lead different polarity degrees. For instance, two rules are tagged negatively whereas three rules are tagged positively in Table 3. Ordered Logit Regression analysis provides us to choose the significant factors with their degree of the impacts on the opinions. This kind of information enables us to focus on some factors instead of all of them that may not be feasible in terms of time and/or other constraints. Here, we choose the aspect *"Diagnosis"* and analyze its relations with other factorial aspects. To illustrate, some statements including the associated rules to provide more insights on aspect connections are presented as follows:

[**Rule #1**] Whenever patients comment on the *Diagnosis* and *Helpfulness* aspects together, they are likely to comment on the *Concern* aspect of the doctor.
◦ (positive) *"Excellent Doctor - diagnosed my cancer and helped me get through it. He is very caring and compassionate."*

[**Rule #2**] Whenever patients comment on the *Listener* and *Diagnosis* aspects together, they are likely to comment on the *Knowledge* aspect of the doctor.
◦ (negative) *"Misdiagnosed Hep A sent me home with a Flu diagnosis. Got sicker went back 6 days later was told it was flu again or thyroid. Did not listen to me as an informed patient - did tell him I was travelling in Mexico. Ended up with 3 days in Hospital. Spends little time with patients. Staff changes regularly, lost or did not have knowledge of previous visits. Office not clean. Do not recommend WILL NEVER GO AGAIN"*

[Rule #3] Whenever patients comment on the *Diagnosis* aspect, they are likely to comment together on *Knowledge* and *Concern* aspects of the doctor.

∘ (pos) *"Dr. X is a great doctor, I was recently diagnosed with IBD and was scared and didnt know what to expect, When I met Dr X, he was so nice and reassured me that I will be ok, I really felt like I was being taken care of. He's a doctor that cares about his patients and he is definitely very knowledgeable. I am feeling a lot better and it's thanks to him."*

To sum up, whenever patients comment on *Listener* and *Diagnosis* aspects of the doctor together, they are likely to comment on his *Knowledge*, too. The corresponding relation of aspect triple is negative. But, whenever patients comment on the *Diagnosis* aspect, they are also likely to comment positively on the *Knowledge* and *Concern* aspects of him. So, *Listener* and *Concern* aspects play significant roles on the decisions of patients on the *Diagnosis* aspect. Likewise, in the rule 4, the presence of the aspect *Helpfulness* in reviews is negatively associated with aspects *Kindness* and *Diagnosis*, whereas the aspect *Listener* is positively associated with these aspects in the rule 5.

Connection types aid us to easily interpret the aspect relations. The polarity of an aspect alone can be positive but when we analyze it under a semantic rule, this aspect may change the polarity of the rule as negative when it co-occurs with other aspects. Here, the important thing is to find out the factorial aspects that change the polarity degree of the rules, and then analyze their relations with other aspects. To ameliorate the current system, consideration of negative $\rightleftharpoons$ positive semantic rule associations are vital. For this reason, we recommend service providers to choose one of the preferred factorial aspects and analyze its relation with other aspects that present in semantic rules. This information extraction can be used as an effective input to better their services and operations management.

In this study, we find out the answers of the following questions like which aspect-pairs co-occur in the texts, what are their relations, and which aspects have significant impacts upon opinions? We can easily reach a decision on the service provider(s) and/or on their services by choosing preferred one or multiple aspects.

8. Conclusion and future work

This paper illustrates a novel feedback-based recommendation framework for service providers with the objective of presenting them a powerful Decision Support System (DSS) including opinion influencing factors and semantic rules (*i.e.*, discerned relationships between factors). We introduce the *opinion influencing factors* also called as *factorial aspects* (FAs) which refer to aspects having significant impacts upon opinions. The joint analysis of semantic rules and factorial aspects are the key feature of this work. We discuss the full processing pipeline from document collections to topic models to structure learning to rule extraction to improving recommender systems. Thus, we introduce a new perspective on recommender systems. Our proposed framework can be easily implemented to any industries.

As a case study, we choose the healthcare industry and apply our methodology on patients' reviews. We discovered that *Concern* is the most frequently used aspect in reviews, yet one unit change (*e.g.*, pos $\rightarrow$ neg) in the *Diagnosis* aspect has the highest influence on patients' comments. Except the aspect *Punctuality*, all the other aspects are found statistically significant, in other words, the occurrence of these aspects in reviews having significant impacts upon opinions. While the occurrence of some of the aspects have higher impacts on positive reviews than negative ones, for some of them the reverse has happened. To provide feedback, we mainly focus on the occurrences of aspects that have higher impacts on negative reviews than the positive ones. We found that the occurrence of the following aspects: *Diagnosis*, *Kindness* and *Staff* in reviews having higher impacts on the negative opinions than the positive ones. To illustrate, we choose the aspect *Diagnosis* which has the highest impact upon the negative reviews compared to positive ones, and analyze its interactions with other FAs. When we consider triple aspect relations associated with *Diagnosis*, we obtain different polarity degrees. For instance, the polarity degree of the aspect triple <Diagnosis, Knowledge, Listener> is positive, whereas the polarity degree of the aspect triple <Diagnosis, Knowledge, Concern> is negative. Thus, we can deduce that *Listener* and *Concern* aspects play significant roles on the decisions of patients on the *Diagnosis* aspect, and service provider should focus on these aspects to better his service. To interpret the rules, connection types of aspects in related rules should be analyzed. For instance, patients like the doctor if his diagnosis is accurate, then patients are likely to find him knowledgeable and concerning. However, patients do not like the doctor if he is not a good listener and his diagnosis may be inaccurate, then patients are likely to found him not knowledgeable. So, poor listening approach of him coupled with his diagnosis may lead patients' discontentment. To improve his service, he should focus on the associations of aspects in the rules. Limitations of this study are as follows: different topic selection techniques can be applied and their performances can be compared for large datasets and messy reviews. To learn the skeleton and establish the DAG, new algorithms can be implemented and their performances can be compared.

Causal rule analysis with time series and demographic data configuring around a feedback-based recommendation system will be our next research. The answers of the following questions for a future study will be considered: How might the decisions of people change in time? Does the time play a significant role upon opinions? How might demographics including income groups (*e.g.*, low or high) or ethnicity of decision makers influence their concerns and comments on chosen topics?

References

Acid, S., de Campos, L., & Fernández, M. (2013). Score-based methods for learning Markov boundaries by searching in constrained spaces. *Data Mining & Knowledge Discovery*, *26*, 174–212.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics*, *32*, 13–47.

Chapman, W., Fizman, M., Chapman, B., & Haug, P. (2001). A comparison of classification algorithms to automatically identify chest x-ray reports that support pneumonia. *Journal of Biomedical Informatics*, *34*, 4–14.

Colombo, D., Maathuis, M., Kalisch, M., & Richardson, T. (2012). Learning high-dimensional directed acyclic graphs with latent and selection variable. *The Annals of Statistics*, *1*, 294–321.

Dehkharghani, R., Mercan, H., Javeed, A., & Saygın, Y. (2014). Sentimental causal rule discovery from Twitter. *Expert Systems with Applications*, *41*, 4950–4958.

Gámez, J., Mateo, J., & Puerta, J. (2011). Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood. *Data Mining & Knowledge Discovery*, *22*, 106–148.

Grün, B., & Hornik, K. (2011). Topicmodels: an R package for fitting topic models. journal of statistical software. *Journal of Statistical Software*, *40*, 1–30.

Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 50–57).

Huang, Z., Lu, X., Duan, H., & Zhao, C. (2012). Collaboration-based medical knowledge recommendation. *Artificial Intelligence in Medicine*, *55*, 13–24.

Kim, H., Castellanos, M., Hsu, M., Zhai, C., Rietz, T., & Diermeier, D. (2013). Mining causal topics in text data: Iterative topic modeling with time series feedback. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management* (pp. 885–890).

Larrañaga, P., Poza, M., Yurramendi, Y., Murga, R., & Kuijpers, C. (1996). Structure leaning of Bayesian networks by genetic algorithms: A performance analysis of control parameters. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, *18*, 912–926.

Li, D., Shuai, X., Sun, G., Tang, J., Ding, Y., & Luo, Z. (2012). Mining topic-level opinion influence in microblog. In *Proceedings of the 21st ACM International Conference on Information & Knowledge Management* (pp. 1562–1566).

Li, J., & Shi, H. (2007). Knowledge discovery from observational data for process control using causal Bayesian networks. *IIE Transactions*, *39*, 681–690.

Lu, Y., Mei, Q., & Zhai, C. (2011). Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA. *Information Retrieval*, *14*, 178–203.

Mahadevan, S., Zhang, R., & Smith, N. (2001). Bayesian networks for system reliability reassessment. *Structural Safety*, *23*, 231–251.

Mani, S., & Cooper, G. (2004). Causal discovery using a Bayesian local causal discovery algorithm. *MEDINFO*, *11*, 731–735.

Margaritis, D., & Thrun, S. (2000). Bayesian network induction via local neighborhoods. *In Advances in Neural Information Processing Systems 12*, (pp. 505–511).

Miller, G. (1995). Wordnet: A lexical database for English. *Communications of the ACM*, *38*, 39–41.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1–135.

Paul, M., & Dredze, M. (2015). SPRITE: Generalizing topic models with structured priors. *Transactions of the Association of Computational Linguistics*, *3*, 43–57.

Paul, M., Wallace, B., & Dredze, M. (2013). What affects patient dissatisfaction? Analyzing online doctor ratings with a joint topic-sentiment model. In *AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI*.

Pearl, J. (2000). Introduction to probabilities, graphs, and causal models. In *Causality: Models, reasoning, and inference* (pp. 1–40). Cambridge University Press. (2nd ed.).

Salmon, W. (1980). Probabilistic causality. *Pacific Philosophical Quarterly*, *61*, 51–64.

Sanchez-Bocanegra, C., Sanchez-Laguna, F., & Sevillano, J. (2015). Introduction on health recommender systems. *Data Mining in Clinical Medicine*, *1246*, 131–146.

Schlüter, F. (2014). A survey on independence-based Markov networks learning. *Artificial Intelligence Review*, *42*, 1069–1093.

Scutari, M. (2009). Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software*, *35*, 1–22.

Settas, D., Cerone, A., & Fenz, S. (2012). Enhancing ontology-based antipattern detection using Bayesian networks. *Expert Systems with Applications*, *39*, 9041–9053.

Su, C., Andrew, A., Karagas, M., & Borsuk, M. (2013). Using Bayesian networks to discover relations between genes, environment, and disease. *BioData Mining*, *6*, 1–21.

Tsamardinos, I., Aliferis, C., & Statnikov, A. (2003). Algorithms for large scale Markov blanket discovery. In *16th International FLAIRS Conference* (pp. 376–380).

Tsamardinos, I., Brown, L., & Aliferis, C. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, *65*, 31–78.

Villanueva, D., González-Carrasco, I., López-Cuadrado, J., & Lado, N. (2016). SMORE: Towards a semantic modeling for knowledge representation on social media. *Science of Computer Programming*, *121*, 16–33.

Wiesner, M., & Pfeifer, D. (2014). Health recommender systems: concepts, requirements, technical basics and challanges. *International Journal of Environmental Research & Public Health*, *11*, 2580–2607.

Zhang, J., Wang, Y., & Vassileva, J. (2013). SocConnect: a personalized social network aggregator and recommender. *Information Processing* & *Management*, *49*, 721–737.

Zhang, N., & Poole, D. (1996). Exploiting causal independence in Bayesian network inference. *Journal of Artificial Intelligence Research*, *5*, 301–328.

Zhao, D., & Weng, C. (2011). Combining PubMed knowledge and EHR data to develop a weighted Bayesian network for pancreatic cancer prediction. *Journal of Bayesian Informatics*, *44*, 859–868.