The Report Committee for Mariia Shovkun
Certifies that this is the approved version of the following Report:

Manual and Automatic Analysis of Patients Values and Preferences Using Seton HCAHPS Surveys

APPROVED BY

SUPERVISING COMMITTEE:

Kenneth R. Fleischmann, Supervisor

Bo Xie

Manual and Automatic Analysis of Patients Values and Preferences Using Seton HCAHPS Surveys

by

Mariia Shovkun

REPORT

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN INFORMATION STUDIES

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2018

Acknowledgments

Prima facea, I would like to express the deepest appreciation to my adviser, Professor KEN FLEISCHMANN for his guidance, advising, and the provided opportunity to participate in an interesting study. Without his supervision and persistent efforts, this report would not have been possible.

I would like to extend my deep gratitude to Professor Bo Xie. Her comments and feedback on my work have been of immense help.

Finally, I take this opportunity to express the profound gratitude to my beloved husband Igor, as well as my parents for the support they provided over the years.

Manual and Automatic Analysis of Patients Values and Preferences Using Seton HCAHPS Surveys

Mariia Shovkun, M.S.Info.Stds.
The University of Texas at Austin, 2018

Supervisor: Kenneth R. Fleischmann

Understanding the values and preferences of patients is a key to patient-centered care. This study aims to identify these values and preferences by analyzing the data provided by The Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). The data is presented as a free-form survey of patients of a variety of hospitals in the United States. The survey was conducted via telephone and concluded with free-form comments that provide a unique way to understand positive and negative sides of health care services. We use topic modeling to automatically identify the crucial aspects of patient-centered care as well as obtain the most high- and low-rated traits of the healthcare system. In particular, we use Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA) to identify the extent to which the topics correspond to values and preferences. In this study we present (1) an approach to manual and automatic evaluation of patients values and preferences and (2) a list of the aspects of the modern healthcare system that are

crucial to its patients. This list can be used by hospitals in order to improve the quality of services they provide. Our findings have important implications for patient satisfaction, patient-provider relationships, and, ultimately, health outcomes.

Table of Contents

List of Tables

Chapter 1

Introduction

1.1 Models for patient-doctor interaction

In first theories about relationship between patient and the doctor the last one was the key element of the treatment. According to Parsons, the fact that patient asked for the help of professionals means rigorous realization of physicians recommendations by the patient [52]. The philosophy of patient and physician roles in healthcare started widely spread developing. For example, the revolutionary concept of patient-doctor relationship was claimed by Emanuel and Emanuel [17]. The model was based on four main components: the physicians role, physicians obligations, patients values, and patient's autonomy. A variety of specialists noticed that approach when the physician main purpose to cure a patient if possible or at least improve his physical condition and prevent disease development does not reflect the patients relation to the doctors opinion. Some authors emphasize that patients treatment satisfaction depends on individual characteristics as well, for example, personal relation to the disease, habits, behavioral features, etc. At the same time, a set of previous studies proved the importance of attentive and respectful communication with patients to the overall patients satisfaction [47, 27]. Great contribution to the overall healthcare improvement was made by Massachusetts Health Quality Partners and California Healthcare Performance Information System (MHQP). Every year the MHQP measures healthcare quality via the patients survey, in particular they measure items, such as "doctor listens carefully", "doctor talks about goals", "doctor spends enough time". Since 2005 MHQP started reporting the survey results, the most valuable part of the research is that after that the quality of healthcare continuously improves. Previous studies ultimately proved the importance of human values in healthcare.

1.2 Closed and open-ended patient surveys

Conducting patient surveys is one of the standard ways to evaluate the quality of the medical service. These service help to identify the key deficiencies in the healthcare system [55]. Most of the surveys are based on sets of pre-defined questions and require to evaluate the particular characteristics of the provided service on a numerical scale. Although this way of conducting surveys helps to understand which known aspects of the healthcare service are the most important and require improvement, it does not allow to identify the new areas of patients' interests [55]. The composition of comprehensive patients' surveys is still an ongoing research effort [8, 13].

Open-ended (also referred to as free-form surveys) surveys are known to contain critical information that cannot be obtained using closed-ended surveys [7, 36, 54]. Although rarely used, this survey type has been recommended for over two decades [42].

1.3 Natural language processing

Ever-growing amounts of data obtained as the patients evaluation of the clinical service are impossible to process individually. As a result, natural-language processing (NLP) began as a science. The first attempts to apply computing to process free text used complex combinations of hand-written rules [33, 31, 37, 28]. Later NLP based on statistical algorithms was introduced and provided more accurate and robust methods for text analysis [34, 43]. The older techniques based on hand-written rules were discarded due to inability to extract semantics from text, poor handling of ungrammatical texts, and the absence of robustness and rigorousness [49]. The modern NLP analysis consists of data preparation and statistical modeling [29]. Nadkarni et al. described the main low-level subproblems in data preparation for medical NLP: tokenization, part-of-speech tagging, morphological decomposition, shallow parsing (chunking), and problem-specific segmentation [49]. Tokenization is the process of separating a document into individual units (words) [50, 9]. In the English language the problem of tokenization is compicated by physical quantities (10 mg/ml), hyphenated compounds (high-risk, long-term), and contractions (arent, isnt), etc.

Lemmatization is the process of grouping various forms and derivatives of a word or set phrases [20, 39, 48]. Various lemmatization solutions usually comprise morphological segmantation (grouping the forms of a word, e.g. doctor and doctors), part-of-speech tagging (e.g. sympathize and sympathetic), shallow parsing (set phrases, e.g., doctor of medicine, intensive care), and stemming (grouping words with the same root) [40, 5, 53, 46, 23, 26, 64]. Significant attention has received NLP using support vector machines (SCV) [12, 41, 57]. This type of statistical learing has been recognized as effective means for classification problems, such as sentiment analysis, text classification, and entity recognition [51, 30, 56]. Sequence models, such as Hidden Markov models (HMM), Conditional Random Fields (CRF), Levenshtein distance models, have been found useful for part-of-speech tagging, inference, identifying errors and plagiarism [59, 61, 25]. Another information-extraction method widely used for NLP is N-grams [60, 6]. Although relatively old, N-gram analysis is commonly used for auto-completion, spelling correction, and semantic analysis in modern NLP [14, 45, 11, 58].

1.4 Topic modeling

Topic modeling is commonly used to identify hidden semantic structures in texts [1]. In this area of text analysis, such methods as Latent Dirichlet allocation (LDA), latent semantic indexing (LSI), also known as Latent semantic analysis, (LSA) has received wide recognition, for they reduce the problem dimensionality, thus, simplifying the analysis [4, 35, 44]. Additionally, parallel implementations of these methods have been known to scale well for many processors [66].

LDA and LSA have been used to solve a variety of problem, including categorizing data for filtering and sentiment analysis, among others [3, 24, 38, 63].

1.5 NLP in medical sciences

Using NLP for analyzing medical surveys is currently an active area of research. Doing-Harris et al. Used a Nave Bayes classifier in order to perform topic modeling on positive and negative patients responses [15]. Their results concluded the correlation between the patients sentiment with the personal characteristics of the medical staff. Maramba et al. used the Keyword in Context approach in conjunction with logistic regression to identified the context in which the key positive and negative words in patients free-text reviews [Maramba et al, 2015]. Their work concludes the ability of NLP to distinguish between positive and negative patients responses and identify the context of the negative service evaluation. One of the limitations of their work is the absence of a specialized dictionary that allows a targeted lemmatization of this particular subject. Greaves et al. demonstrated a relatively high capability of a priori classifiers to extract the estimates of the hospital cleanliness, treatment with dignity, and overall satisfaction [19].

The study "A Hospital Recommendation System Based on Patient Satisfaction Survey" explores the data provided by HCAHPS survey [32]. The survey contains close end questions about patients experience. The authors found the relation between overall patients satisfaction, satisfaction of particular service, for example, nurse or doctor care, symptoms info, help after discharge, etc. and patients characteristics (gender, age, nationality, etc.). They used analysis of variance to predict possible challenges with patients with particular list of parameters. ANOVA is a supervised machine learning technique which allows to see the relation between factors, which is impossible in our case.

In this study we offer patient-centered approach to identify values from the patients perspectives, instead of using social and medical sciences categories. For this purpose we analyzed data provided by Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) [32]. Chapter 2 presents the description of the data set and the methodology for the analysis conducted in the study. Chapter 3 presents the results of the analysis that consists of statistical and manual data processing. We discuss certain limitations and extensions of our work in Chapter 4. Finally, we outline the main conclusions of this work in Chapter 5.

Chapter 2

Methods

2.1 HCAHPS free-form survey

HCAHPS free-form survey differs from other surveys because it contains comments from patients in a free form instead of multiple-choice questionnaire [32]. The participants (patients) were asked only two questions. The first one was to describe anything outstanding during a patient's hospital stay. The second question was: "What could have been done to improve (your/your family member's) hospital stay?"

For the privacy reasons, all hospital team member names as well as the names of the hospitals were removed from the data set. The data set contains information regarding the span of the patients stay and the date of the survey. The information about the patients, such as their age or gender is unknown.

2.2 Data processing

As described above, the survey participants had ts answer two questions that were designed to separated positive and negative patients experiences. Due to this separation, we performed our analysis on two sets of data: the first containing the positive patients reviews, and the second containing the negative experience reviews. The overall workflow of data processing is the same for both data sets, but certain parts of it differ due to the nature of the data sets.

In order to find the most typical patients preferences and values that are crucial to the overall patients satisfaction, we used a combination of manual and automatic techniques. We first describe the automatic methods used in the study.

2.2.1 Step 1. Data pre-processing.

The first step of the algorithm is fully automated. We start with removing punctuation, double and trailing whitespaces, and recapitalizing the text. This step is necessary due to the ubiquity of the punctuation symbols and its irrelevance for the analysis and in order to eliminate the difference between the words at the beginning and middle of sentences.

We next tokenized the string of sentences in to the list of words using NLTK library [2].

Third, we removed the stop words, i.e., the words which does not provide any relative information about the topic, for example "for", "I", "is", and "was", etc. This part is important because these words with the highest frequency influence the words dimensionality without providing information. We used the NLTK corpus stop words dictionary to perform this task.

Fourth, we lemmatized the words, which means that each word with same root was transformed into the same form. For example, in the set “nurses”, “nurse”, “nursery”, “nursing” the NLTK lemmatize function replaces all these words with the word “nurses”.

2.2.2 Step 2. Manual data pre-processing.

First, we removed non-informative words from the list of tokenized words. Many of those words were personal or names (privacy reasons), hospital names, or the titles of specific medicaments. We also removed all the words with the frequency in the text equal to one as they do not contribute to the statistical distributions but slow down the performance of the analysis. Then, we removed the words that are non-informative according to our subjective opinion. This procedure consists in filtering out the most frequent uninformative words from the set of all the words in the text. Thus, we obtained a problem-specific list of stop-words. This comprised the numbers, the months names, days-of-week, and personal names. Additionally, such words as good, amazing, and fantastic were considered uninformative because they occurred only in the positive reviews. The stop words list also included genders, nationalities and the words that characterize personal appearance, such as black, blond, or asian since these words should not influence the quality of healthcare services. This composition of the stop word list was iterative: we updated the list during the analysis as long as they appeared in the output of the statistical analysis described in Step 3.

Second, we created a list of synonyms in order to classify words with similar meanings as the same words. This process has the same objective as lemmatization and as well helps to reduce the number of unique words in the data set. Taking into account the specifics of hospital environment, we assumed that the words in such pairs as "friendly" and "nice", "fast" and "quick", and "physician" and "doctor". have similar meanings. Similarly to the list of stop words, the composition of the list of synonyms was iterative and was performed as long as the statistical analysis yielded synonyms.

2.2.3 Step 3. Statistical analysis

We opt to use three statistical approaches that are widely-used for topic modeling: Latent Dirichlet Allocation (LDA), Latent Semantic analysis (LSA), and Hierarchical Dirichlet Process (HDP) [16, 4, 62].

Latent Dirichlet Allocation. LDA assumes that each document (review) is a set of different topics, and the prior distribution is the Dirichlet one, as well as the words in the topic are Dirichlet distributed.

Latent Semantic Analysis. The LSA model constructs the matrix with the document number in the column and the words from each document in a row. From this matrix it calculates the similarity between documents and between words by calculating the product of rows and columns respectively.

Hierarchical Dirichlet Process. The method is, in fact, an extension of LDA. The main difference between the models is that in HDP the number of topics is not known a priori. HDP model generates the number of topics by Dirichlet process, which add one more level of modeling process.

The idea of all three methods is to consider the documents as a set of topics distributed in a certain proportion. Similarly, each topic is considered to be a set of key words. The purpose of each method is to find the optimal composition of the key words within each topic.

One limitation of using these methods for free-form survey analysis the length of the text. Typically, these techniques are applied to documents that comprise a variety of topics. In particular, these methods refer a document to the specific theme. For example, an article that contains the words cat, dog, and elephant is probably related to the animal theme. In the same way, the search engines use sentiment analysis to allocate related web-source according to the search request. In other words, such type a techniques allow to reduce dimension of set of words to the lower dimension of topics. In the case of free-form surveys, each document has only one top level theme: patients experience at the hospital. Nevertheless, the methods described above are still capable to find the key points in the treatment experience. The aforementioned techniques are based on the words relation within the one review. The stronger the relation between the words in the same review is, the higher is the probability that these words are related to a particular topic.

LDA, LSA, and HDP methods yield a finite number of words per topic as the output. The number of words per topic is user-chosen for LSA and LDA methods, and defined automatically in the HDP method. Each word within the topic has a weight, the higher the weight, the more important the word for particular topic. Due to the statistical basis of these methods, their results are extremely sensitive to the words frequency in the documents. Obviously, such words as “I”, “the”, and “me”, appear more frequently in the survey than the words with the semantics more meaningful for the analysis (in terms of topic extraction).

Therefore, an essential part of this work was to maximize the frequency of the informative words in the documents. Additionally, because such short documents as free-form answers may contain a variety of synonyms, we performed manual data processing to reduce dimensionality in the analysis.

2.2.4 Step 4. Contextual analysis of the dictionary

We finally perform a manual analysis of the synonyms dictionary obtained after iterating over Steps 2 and 3. The words in the dictionary are represented by a set of key-value pairs, where the words are the keys and the values are categories that these words belong to. We manually analyzed the semantic meaning of each category in the dictionary and classify them into several top-level categories as shown in Section 3.4. Special attention in our analysis has been given to adjectives. As exemplified in Chapter 3, whereas most words can be uniquely categorized per se, many adjectives are impossible to interpret without context. Therefore, our top-level and medium-level categorization was performed with an account for the context of ambiguous adjectives.

Chapter 3

Results

3.1 Analysis with no case-specific preprocessing

We first present the results of the statistical analysis without problem-specific processing (performing Step 1 and Step 3 of the algorithm only). Table 1 shows five arbitrarily-selected topics with the corresponding lists of words obtained by applying LDA model with 20 topics and 7 words per topic to the positive reviews data set.

As shown in Table 3.1, the word nurse presents in all five categories. As discussed above, this is due to the high sensitivity of the statistical methods to the frequency of the words in the set of documents. The results contain many words that are irrelevant for the semantic analysis, such as “ever”, “really”, and “away”. The results also indicate that numbers (sixth, two, one) have high frequency in the set of responses. The word name that is a substitute for personal names is also frequently encountered in the text despite its irrelevance for the analysis. Finally, the topics selected by the model do not constitute semantically-holistic groups.

Similar results were observed when applying LSI and HDP model to the positive and negative datasets without problem-specific stopwords and dictionaries. The reason of the impossibility to interpret these results is twofold: (1) an abundance of the words that are semantically irrelevant and (2) high ratio of the number of topics per the number of words in a topic.

Topic #1	Topic #2	Topic #7	Topic #12	Topic #18
overall	would	stay	nurse	feel
care	nurse	nurse	name	right
other	amazing	level	two	comfortable
experience	come	hospital	named	made
hour	name	best	one	away
nurse	explain	care	lactation	nurse
doctor	came	health	really	also

Table 3.1: Topics identified with LDA model without problem-specific preprocessing for the data set of positive reviews. Abundant semantically-irrelevant words, such as other, hour, and best impede topic modeling with LDA.

3.2 Analysis after eliminating problem-specific stop words

We next present the results of the statistical analysis after discarding problem-specific stop words that hindered the previous analysis. Same as in the previous Section, we performed an LDA run with 20 topic and 7 words per topic. Five arbitrary-selected topics obtained by the model application to the positive data set are shown in Table 3.2.

Topics 4, 7, and 17 are clearly related to the responses that expressed satisfaction with the professional and personal attitude of nurses and doctors to the patients. Topic 11 describes treatement cases that involved surgery. Topic 19 concerns the cases when the patients were pleased with the agile nurse response to the call button in neonatal (newborn) intensive care units

(NICU). Discarding problem-specific stop words results in the better performance of the LDA model than in the case with no data pre-processing. The results, however, contain a variety of semantically-similar topics (such as topics 4, 7, and 17) due to the models inability to group the words by meaning. Due to the same reason, some topics contain synonyms within them, e.g., Topic 52 comprises the words “courteous” and “supportive”, which are both related to the attitude of the staff. Combining the words “surgeon”, “doctor” and “anesthesiologist” in Topic 17 could also cause the appearance of new words that would improve the context of this topic. Therefore replacing such words, as care, help, compassion, and smiling, with a single category can force the model to group semantically-similar topics, as well as to reduce the dimensionality of the problem (that results in appearance of new topics).

Topic #4	Topic #7	Topic #11	Topic #17	Topic #19
help	caring	surgeon	helpful	nicu
nurse	nurse	nurse	nurse	button
provided	compassion	notch	staff	nurse
lady	care	doctor	courteous	quick
sweet	pleased	anesthesiologist	supportive	pushed
care	support	surgery	care	care
responded	smiling	helpful	doctor	brought

Table 3.2: Topics identified with LDA model with problem-specific stop words but without synonyms dictionary for the data set of positive reviews. The abundance of synonyms results in (1) the appearance of semantically-identical topics and (2) multiple synonyms per topic.

3.3 Analysis after eliminating stop-words and applying problem-specific dictionaries

In this section we present the results of statistical processing after eliminating problem-specific stop-words and combining synonyms within the context of a patient survey. Additionally, we compare the results obtained from LDA, LSA, and HDP models. We performed Step 3 of our algorithm with 20 topics and 7 words per topic.

Table 3.3 summarizes the results of the statistical analysis performed on the data set of positive review.

- The results obtained with LDA model are shown in Table 3.3a. The topics selected by the model allow to identify their semantic meaning. For example, Topic 1 may describe a group of patients satisfied with the attitude of the personnel provided in the intensive care; these patients were also pleased because they were able to rest without being constantly disturbed. Less specific representation than in Table 3.2. Topic 2 probably corresponds to a set of reviews that pointed out the responsiveness of the staff at nights. Topics 4 and 5 describe the patients that were particularly satisfied with food and were in labor, respectively. It is important to point out that the semantic meaning of the automatically-obtained topics is difficult to interpret in some cases. This is due to the generalization introduced by the synonyms dictionary. Such words as "surgeon" or "anesthesiologist" most likely have the context of a surgical operation, which is lost when replacing these words with a generic word "doctor".

Thus, using LDA for topic modeling in context of patients' surveys yields semantically-interpretable results.

- Table 3.3b presents the results of analysis performed with LSA. Similarly to LDA, LSA-analysis selects topics that may describe particular groups of patients. Some of these topics, such as Topic 1 and 3 are similar to those identified by the LDA algorithm. Topics 2, 4 and 5 reveal new points of interests: the attitude to the patient's family, hospital food, family accommodation, and pain management. A notable difference between the results obtained with LDA and LSA is that in the latter one can observed more semantic points per topic. This result may indicate that LSA algorithm may be more suitable for the documents with high numbers of topics per document, such as free-form surveys.

- The topics with the corresponding words obtained by the HDP analysis are shown in Table 3.3c. In some cases (e.g. Topic 1), HDP produces topics that represent the overall patient's satisfaction with the work of the staff. More frequently, however, HDP grouped words with a very low frequency in the text. Many of these words were given and family names, that we failed to eliminate with the list of problem-specific stop words. Another group of low-frequency words observed in HDP results consisted of misspelled words ("asain", "docotors") and contractions (e.g., "med"). Overall, the majority of the topics identified by HDP are irrelevant to our analysis. Therefore, although HDP might produce meaningful topics when applied to free-form surveys, this tool performs poorly on this type of analysis. Nevertheless, HDP may be a good means for data preprocessing, e.g., to eliminate typos and personal names from the text.

Topic #1	Topic #2	Topic #3	Topic #4	Topic #5
		(a) LDA		
staff	staff	treat	food	staff
care	night	staff	staff	attention
rest	attention	fast	satisfaction	care
attention	head	attention	light	information
patience	care	area	expectation	labor
expectation	quick	arrival	care	room
believe	diligent	insurance	nurse	labor
		(b) LSA		
staff	attention	information	treat	check
visit	staff	care	night	visit
response	care	labor	comfortable	listen
respect	information	quality	patient	patient
care	labor	attention	family	care
quick	family	area	respect	call
helpful	room	intensive	food	pain
		(c) HDP		
staff	staff	staff	staff	insert
attention	attention	attention	patient	blown
care	care	care	name #2	hospital
labor	late	agreeable	explaing	name #5
information	asain	name #1	name #3	disorder
room	spotless	slow	name #4	name #6
intensive	bald	docotors	chubby	med

Table 3.3: Topics identified with LDA (a), LSA (b), and HDP (c) models on the data set of positive review. (a) LDA model selects the generic topics of interest. (b) LSA selects generic topics and performs similarly to LDA. (c) HDP yields many low-frequency semantically-irrelevant words, such as personal names. The topics identified by HDP carry ambiguous semantic meaning.

We now discuss the results of the statistical analysis performed on the data set of negative reviews. Table 3.3 presents five arbitrarily-selected topics given by LDA, LSA, and HDP analysis with 20 topics and 7 words per topic.

- The topics selected by LDA are shown in 3.3a. Each topic given by the model permits to infer the context of the reviews corresponding to this topic. For example, Topic 1 is probably related to a poor selection of food and issues related to serving food. Topic 3 refers to the room, and the work of the TV set in particular. Some topics, e.g., Topic 2, contains a set of keywords that is very similar to topics inferred from the data set of positive reviews. Overall, the topics selected from the negative reviews are more specific than those selected from the positive reviews. This is due to the fact that the patients tend to generalize their overall experience when they are satisfied

- We show the topics selected by LSA analysis in Table 3.3b. Similarly to LDA, the results given by LSA may be semantically related to a particular context. For example, Topic 3 comprises the complaints that are likely related to the accommodation during labor and poor or absent painkiller drugs. Topic 5 is likely to describe the complaints about the long wait for the call button response at night.

- The results of the HDP analysis on the set of negative reviews (3.3c) consist of many low-frequency words, such as medical terms, personal names, and misspelled words. Therefore, HDP performs on the set of negative reviews in the same way as on the set of positive reviews. The topics selected by HDP forbid to determine their context.

Topic #1	Topic #2	Topic #3	Topic #4	Topic #5
		(a) LDA		
diet	clean	television	provide	noise
helpful	staff	channel	nicu	night
variety	visit	easy	family	allergic
late	hot	accommodation	option	loud
staff	bathroom	clean	labor	room
serve	housekeeping	turn	room	hall
improve	attention	broken	plan	incident
		(b) LSA		
staff	pain	medication	night	call
service	medication	pain	information	visit
choice	attention	room	family	wait
diet	staff	accommodation	medication	long
quality	room	button	sleep	button
cold	uncomfortable	labor	attention	information
taste	medicine	staff	check	night
		(c) HDP		
staff	staff	staff	staff	staff
indisposed	sense	yucky	tended	name #1
term	embulization	chemical	defeated	mri
believed	pig	intruded	prescritption	gettging
administration	admin	food	refurbishing	rubber
grilled	codeine	biggest	suggest	childbirth
uro-gynecologist	wai	degree	pigeon	accordance
preventing	remind	diets	foam	room

Table 3.4: Topics identified with LDA (a), LSA (b), and HDP (c) models on the data set of negative review. (a) LDA analysis selects semantically-meaningful topics. (b) LSA selects performs similarly to LDA. (c) HDP yields many low-frequency semantically-irrelevant words, such as personal names, and typos.

3.4 Categories selected from the manual analysis

We now present the results of the manual contextual analysis of the obtained problem-specific dictionary described in Step 4. The results of the analysis are presented in Tables 3.4.1 and 3.4.2 for the data sets of positive and negative reviews, respectively. For each of these data sets we identified three top categories titled "Hospital Team", "Treatment", and "Accommodation". Each top category contained several medium-level categories, as shown below.

3.4.1 Points of interest identified in positive reviews

In this section we present the key points of interest retrieved from the data set of positive reviews. These results are presented in Table 3.4.1 and explained in detail below.

1. **Hospital team** We found two medium-level topics for this top-level category.

    (a) **Staff**. These points of interest that are common for both doctors and nurses.

    - **Respect.** Most of the patients noticed the respectful relation from the doctors and especially from the nurses. Several patient mentioned respectful relation from the overall hospital team to their privacy.
    - **Friendliness.** the overwhelming majority of patients used

words “friendly”, “nice”, “kind”, and “courteous” to characterize the staff.

- **Overall care.** Some part of the reviews did not include any detailed description of particular experiences. A significant number of patients evaluated their experience as “good”, “excellent”, and “outstanding”.
- **Professional/knowledgeable.** A substantial number of the patients emphasized the level of professionalism and knowledge of the hospital workers. Some of the participants noticed the overall professional degree, while the others specified the professionalism of particular nurses or doctors.
- **Visits.** As discussed in Section 4.2, the word “visit” was used in several contexts. Most of the patients appreciated the visits from doctors and nurses.
- **Information.** Some patients acknowledged in their reviews that the doctors and nurses provided valuable information about the patients’ diagnosis and the upcoming treatment. Additionally, the staff performed well in answering the questions.
- **Information to family.** Some reviews expressed satisfaction of the information provided by the staff to the family members.
- **Attention and helpfulness.** One of the most frequent topic of the reviews was related to the attention and care provided by the hospital employees.

- **Professionalism.** The professionalism of the staff is a popular topic in the reviews. In some reviews this term was related to the attitude of the employees, in others it characterized the treatment received. In the majority of reviews the use of this term was unspecified.
- **Personal approach.** Some patients were pleased with the interest to the their needs from the staff.
- **Check frequency.** Some reviews emphasized the importance of how often the nurses and doctors checked on the patients needs, conditions or wishes.

(b) **Doctors**. The topics that are related to doctors only is as follows:

- **Medication change.** Several patients noticed that because of their complaints about the current types of medication, particularly pain relief drugs, the doctors changed the medication.
- **Communication with nurses.** A group of patients was pleased with the effective interaction between the doctors and nurses that concerned the information about the patient's condition and needs.

(c) **Nurses**.

- **Pain management.** This term was frequently used in context of effective medication and the change of medication when ineffective.

- **Responsiveness.** Many patients were satisfied with the agility of the nurse reaction on the call button. Some patients emphasized the response time during the night time.
- **Family needs.** The nurses were helpful with providing information, meals, accommodation, and assisting with other family needs.
- **Schedule.** Nurses adjusted their schedule to the patients regime. The medication schedule were comfortable for the patients as well.
- **Listening.** This topic is related to the ability of the nurses to listen and discuss the patients needs. According to the reviews, this quality allowed the patients to stay calm while in a hospital.
- **Humour.** Staff's sense of humor was highlighted by a group of patients.

2. **Treatment.** We identified three medium-level categories for this top-level category.

   (a) **Department.** This topic includes many names of medical procedures. We omit the description of some topics in this category as they are self-explanatory.

   - **Staff.** Since the medical personnel and treating the patient are inextricably linked, we include this topic in this category.

- **Check-in/out process.** For some patients highlighted the satisfaction with the ease and duration of the check-in and check-out procedures.
- **Labor and delivery.** Some patients appreciated the organization of the delivery process, which allowed them to get a satisfactory amount of sleep.
- **Atmosphere.** A group of patience appreciated the environment in the hospital.

(b) **Procedure.** This category is related to surgical operations.

- **Calm.** The personnel made everithing to the patient to ensure the patient is in safety.
- **Informative to patient/family** The staff explained the necessary details of the surgery to the patient or family.

(c) **Treatment.** This category comprises the topics of pain management and the overall duration of their hospital stay. The reviews that mentioned pain management emphasized the effectiveness and change of drugs.

3. **Accomodation.** We identified four medium-level categories for this top-level category.

(a) **Room.** Many patience appreciated the availability of accommodation for their family members. Other patients pointed out the work of the nurse call button and the cleanness of the rooms.

(b) **Food.** Some reviews mentioned the high quality of food served at the hospital. Other patients enjoyed the food delivery to their rooms. Interestingly, this is a second context of this word in medical service.

(c) **Administration.** This topics includes the reviews of administrative tasks, such as billing information and the effectiveness of check-in and check-out procedures.

(d) **Other.** This topic includes a list of miscellaneous sub-topics related to accommodation. It comprises family needs, the availability of a chapel and a preacher. Finally, some people also pointed out the ease of transportation to the hospital from their places of residence.

Hospital team	Doctor	attentive, frequent visits, treatment knowledgeable, medication change, respectful, informative for patient, informative for family, communication between doctor and nurses
	Nurse	attentive, pain management, availability, informative/family, informative/patient, respect to family, attentive to family, medication schedule, overall care, quick night response, listening, regular checking, responsiveness
Treatment	Department	emergency room, staff, intensive care informative, check-in process, quick admission, labor and delivery, breastfeed help, nurses delivery schedule, lactation consultation, surgery, equipment, atmosphere, NICU, overall care
	Procedure	surgery, calm, informative/patient informative/family, quick
	Treatment	pain management, duration
Accommodation	Room	room for family, clean, call button
	Food	delivery, quality
	Administration	billing help, quick check-in/out
	Other	family needs, preacher, transportation

Table 3.5: Top- and medium-level categories with the corresponding keywords obtained from the positive reviews

3.4.2 Points of interest identified in negative reviews

The high-level and middle-level topics identified from the negative reviews are listed in Table 3.4.2. Comments and detailed explanation of the specific keywords are as follows:

1. **Hospital team** We found two medium-level topics for this top-level category. We first describe several aspects that were common for doctors and nurses.

   (a) **Staff**. These points of interest that are common for both doctors and nurses.

   - **Information.** A group of patience has been concerned about the lack of information obtained from the staff.
   - **Communication between staff members.** Some patients complained about the lack of communication between various hospital departments. Several patients faced miscommunication between doctors and nurses.
   - **Waiting time.** Many complaints were related to the amount of time that patients had to wait for the doctor and to the nurse response time to the call button.
   - **Visit frequency** A group of patients was disappointed by the low visit frequency by doctors and nurses.
   - **Rudeness.** Some patients faced rudeness from both doctors and nurses.

(b) **Doctors.**

- **Attendance.** A variety of complaints were about the doctor's unavailability. Other reviews emphasize that the doctors did not provide sufficient information about the treatment
- **Patient involvement.** Some patients complained that they were not involved in the process of treatment plan development.

(c) **Nurses.**

- **Overworked.** A variety of complaints mentioned that nurses were often performed their work poorly due to being "overworked".
- **Perfume.** Some survey participants complained about the smell (e.g., "too much" or "too strong") of the nurses' perfume.
- **Noise.** A number of complaints were about the noise induced by the work of nurses, particularly during the night time.

2. **Treatment.** We selected three sub-categories for this top-level category.

(a) **Department.** The this comprised the following important keywords.

- **Wait time.** A number of patients complained about the time they had to wait before particular procedures and tests.
- **Transportation.** This topic regards the complaints about the system transportation between hospital units.

- **Dismissiveness.** This topic comprises the complaints about patients' wishes and suggestions being disregarded.

(b) **Treatment.**

- **Consultation.** This topic includes the absence or quality of consultation received from the staff.
- **Birth plan.** Similarly to **Dismissiveness**, the patients complained about their wishes not taken into account.
- **Lack of information provided to family members.**

(c) **Procedure.**

- **Schedule.** A number of patients expressed dissatisfaction with the schedule of procedures that resulted in a lack of sleep for patients.
- **Pain management.** Some reviews complained about the refusal to a request of painkillers, as well as the effectiveness of these drugs.

3. **Accommodation.** This top-level topic consists of four medium-level topics.

(a) **Building.** This category comprised several important key points.

- **Call button.** These complaints were about the absence and the duration of response to pressing the nurse call button.
- **Room.** A number of patients complained about dirty linens and floor.

- **Housekeeping.** Some patients were dissatisfied with the low frequency or long wait for housekeeping (cleaning the rooms).
- **Noise.** Reviews that contain this topic describe two situations: the noise induced by the work of the nurses and the noise from construction.
- **Hot water.** This topic comprises complaints about the absence of hot water or about the water being insufficiently hot.
- **Temperature.** Many reviews emphasized that the rooms were either too hot or too cold.
- **Wheelchair.** A number of complaints were about the absence of facilities for the people in the wheelchair.
- **Church.** Some reviewers complained about the absence of hospital chapel.
- **Television.** A group of negative reviews concerned the lack of TV channels.

(b) **Food.** A number of complaints were about the hospital food.

- **Special food.** Several patients complained about the absence of diet, diabetic, and gluten-free food.
- **Food for visitors.** Several patients complained that the hospital did not provide the food for their visitors.

(c) **Family.** A number of complaints were about service provided to the family members that stayed with the patient during his stay.

- **Food.** Reviewers complained that the hospital did not provide food for the staying family members.
- **Communication.** Staff members refused to provide information to the family members.

(d) **Administration.** This medium-level category consists of several aspects of the administrative process. The survey participants complained about the

- long check-in and check-out procedures,
- the cost of the provided service,
- issues with billing and insurance, and
- late billing information.

Hospital team	Doctor	lack of answers, long waiting, communication to nurses, rare visits, rude, attendance
	Nurse	rude, judgmental, responsiveness overall care, overworked, attention, perfume, loud, patient involvement
Treatment	Department	emergency, long wait, empathy informativeness, doctor, transportation, labor and delivery, dismissive, delivery schedule, consultation, breast feeding help, long wait, birth plan NICU, informativeness/family access to family, location
	Procedure	informativeness, wait for painkillers
	Treatment	pain management, aftercare, surgery, diagnosis and testing schedule
Accommodation	Building	location, call button, dirty linen, housekeeping, noise, reconstruction, no hot water, temperature, television wheelchair, chapel service, theft
	Food	wait, special food, cold, tasteless, food for visitors, quality, choice
	Family	meal, accommodation, rudeness, communication with staff
	Administration	long check-int/out, cost, billing, late billing information, insurance

Table 3.6: Top- and medium-level categories with the corresponding keywords obtained from the negative reviews

Chapter 4

Discussion

4.1 The impact of the size of the problem-specific dictionary

The results presented in Sections 3.2 and 3.3 indicate that introducing problem-specific dictionaries may result in the loss of context within the topic found by statistical models. The specifics of the obtained results is significantly affected by the size of the problem-specific dictionary used for data pre-processing. Using dictionaries that contain many words yields topics with more generic topics than using those with fewer words. Thus, when composing problem-specific dictionaries one needs to be cautious with the amount of synonyms in the dictionary.

We suggest that the number of words in a dictionary should be limited depending on the task. If the analysis aims to identify the generic areas of the healthcare system that require additional attention for budget management, large dictionaries can be used. If one is interested in finding more intricate and specific aspects of the medical services, small dictionaries are recommended.

4.2 N-gram analysis on middle-level categories

The meaning many words used in the survey depends on the context substantially, e.g, care (intensive care, care for patients), visits (family visits, visits by doctor), and delivery (labor, food delivery). Since some of these words are frequently encountered in the survey, they affect the results of the topic modeling significantly. These words, however, cannot be handled by the problem-specific dictionaries due to the ambiguity in their meaning.

In the analysis presented in Section 3.4, we manually analyzed the context of the words with ambiguous meaning. This part of work can be automated using N-grams. According to previous studies, N-gram analysis can successfully applied to topic modeling [6, 65]. This type of analysis is potentially more suitable to sets of documents with multiple topic per document, such as free-form surveys. Additionally, N-gram analysis can mitigate the issues described in Section 4.1 [10].

4.3 Implications

In this study we presented a method to identify the key points of patient's satisfaction and dissatisfaction. This presented algorithm can be used to find the areas that require special attention in specific hospitals. Although the results obtained in this study are list many topics in both positive and negative categories, we emphasize that they were obtained from a survey conducted in a variety of hospitals. A similar analysis performed on the reviews obtained from a single hospital is likely to yield more distinctive differences between the topics in positive and negative reviews.

We emphasize that the presented topic modeling method is not limited to the patients' surveys only. Several areas, such as web comments related to product reviews or reaction to social events, also are free-form texts with a high number of topics per document. A missing part of our study was automated sentiment analysis, i.e., automatic recognition of positive and negative reviews. When combined with problem-specific stop words and dictionaries, a similar analysis can be performed for topic modeling on areas other than medical surveys.

4.4 Reliability

In the previous section we discussed that the results of our analysis were obtained from reviews held across multiple hospitals and may be inapplicable to the work of a particular hospital. There is, however, another source of found can not be applicable for any patients. Some authors describe the doctors relationships experience with the difficult patients [21, 18, 22]. They found that healthcare result could be worse than it could be because of difficulties in patient-doctor relationship. The problems in relationship can arise from both sides: difficult doctor or difficult patient. That is possible that doctor was not attentive enough, on the other hand a patient required especial attention to his person because behavioral or psychopathology features. James E. Groves identified four types of hateful patients [21]. The first one is Dependent Clingers. This type of patients honestly require extra attention to themselves in a various of available ways, it can be continuously repeated questions or requests to the doctors or nurses, sedatives or any other medication requests, etc. The author notices that this form of attention dependency can be the consequence of psychology problems, at the same time clingers usually are not aggressive, grateful, and their demands are naive and instinctive.

Chapter 5

Conclusion

This report presents an analytical study of the free-form patients' surveys. We used a combination of statistical modeling techniques and manual data processing on the data collected from multiple hospitals in order to identify the key points of patients' satisfaction and complaints. Conclusions drawn from this work are as follows:

- We found and summarized the topics that arise in positive and negative patients' reviews most frequently.
- Patients tend to be more specific when describing negative experiences.
- LDA and LSA models are cable of selecting semantically-meaningful topics from free-form surveys.
- The majority topics identified by HDP model do not constitute semantic units.
- HDP analysis can be used to find misspelled words, personal names, and other low-frequency entities in free-form texts. We recommend to use LDA for data pre-processing.

- Traditional topic-modeling tools, such as LDA, LSA, and HDP, are unable to identify low-level unitary semantic topics in free-form surveys as the latter contain high numbers of topics per document.

- Data pre-processing with problem-specific stop words and dictionaries improves the results of the statistical analysis.

- Using problem-specific dictionaries with high numbers of synonyms may result in the potential lack of context in topics selected by LDA and LSA.

- LSA analysis is capable of selecting more specific topics than LDA at the same size of the dictionary.

- N-gram analysis was proposed as a means select unitary topics within multi-topic short documents.

Bibliography

[1] David Andrzejewski, Xiaojin Zhu, and Mark Craven. Incorporating domain knowledge into topic modeling via dirichlet forest priors. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 25–32. ACM, 2009.

[2] Steven Bird and Edward Loper. Nltk: the natural language toolkit. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, page 31. Association for Computational Linguistics, 2004.

[3] István Bíró, Dávid Siklósi, Jácint Szabó, and András A Benczúr. Linked latent dirichlet allocation in web spam filtering. In *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*, pages 37–40. ACM, 2009.

[4] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. In *Advances in neural information processing systems*, pages 601–608, 2002.

[5] Eric Brill. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational linguistics*, 21(4):543–565, 1995.

[6] Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479, 1992.

[7] Stephen Bruster, Brian Jarman, Nick Bosanquet, Debra Weston, Robert Erens, and Thomas L Delbanco. National survey of hospital patients. *Bmj*, 309(6968):1542–1546, 1994.

[8] Raymond G Carey and Jerry H Seibert. A patient survey system to measure quality improvement: questionnaire reliability and validity. *Medical care*, pages 834–845, 1993.

[9] Tagyoung Chung and Daniel Gildea. Unsupervised tokenization for machine translation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 718–726. Association for Computational Linguistics, 2009.

[10] Marc Damashek. Method of retrieving documents that concern the same topic, May 23 1995. US Patent 5,418,951.

[11] Anh Dang, Abidalrahman Moh'd, Aminul Islam, Rosane Minghim, Michael Smit, and Evangelos Milios. Reddit temporal n-gram corpus and its applications on paraphrase and semantic similarity in social media using a topic-based latent semantic analysis. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3553–3564, 2016.

[12] Hal Daumé III. Support vector machines for natural language processing. *Lecture Notes*, 2004.

[13] Elizabeth A Davies, Mark M Meterko, Martin P Charns, Marjorie E Nealon Seibert, and Paul D Cleary. Factors affecting the use of patient survey data for quality improvement in the veterans health administration. *BMC health services research*, 11(1):334, 2011.

[14] Li Ding, Tim Finin, Anupam Joshi, Rong Pan, R Scott Cost, Yun Peng, Pavan Reddivari, Vishal Doshi, and Joel Sachs. Swoogle: a search and metadata engine for the semantic web. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 652–659. ACM, 2004.

[15] Kristina Doing-Harris, Danielle L Mowery, Chrissy Daniels, Wendy W Chapman, and Mike Conway. Understanding patient satisfaction with received healthcare services: A natural language processing approach. In *AMIA Annual Symposium Proceedings*, volume 2016, page 524. American Medical Informatics Association, 2016.

[16] Susan T Dumais, George W Furnas, Thomas K Landauer, Scott Deerwester, and Richard Harshman. Using latent semantic analysis to improve access to textual information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 281–285. Acm, 1988.

[17] Ezekiel J Emanuel and Linda L Emanuel. Four models of the physician-patient relationship. *Jama*, 267(16):2221–2226, 1992.

[18] TJ Gerrard and JD Riddell. Difficult patients: black holes and secrets. *BMJ: British Medical Journal*, 297(6647):530, 1988.

[19] Felix Greaves, Daniel Ramirez-Cano, Christopher Millett, Ara Darzi, and Liam Donaldson. Machine learning and sentiment analysis of unstructured free-text information about patient experience online. *The Lancet*, 380:S10, 2012.

[20] Nathan Green, Paul Breimyer, Vinay Kumar, and Nagiza F Samatova. Webbanc: Building semantically-rich annotated corpora from web user annotations of minority languages. 2009.

[21] James E Groves et al. Taking care of the hateful patient. *Personality Disorder the Definitive Reader*, pages 52–63, 1978.

[22] Steven R Hahn, Kenneth S Thompson, Thomas A Wills, Virginia Stern, and Nancy S Budner. The difficult doctor-patient relationship: somatization, personality and psychopathology. *Journal of clinical epidemiology*, 47(6):647–657, 1994.

[23] James Hammerton, Miles Osborne, Susan Armstrong, and Walter Daelemans. Introduction to special issue on machine learning approaches to shallow parsing. *Journal of Machine Learning Research*, 2(Mar):551–558, 2002.

[24] Niklas Jakob, Stefan Hagen Weber, Mark Christoph Müller, and Iryna Gurevych. Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 57–64. ACM, 2009.

[25] AJPMP Jayaweera and NGJ Dias. Hidden markov model based part of speech tagger for sinhala language. *arXiv preprint arXiv:1407.2989*, 2014.

[26] Anjali Ganesh Jivani et al. A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl*, 2(6):1930–1938, 2011.

[27] Steven Joffe, M Manocchia, JC Weeks, and PD Cleary. What do patients value in their hospital care? an empirical perspective on autonomy centred bioethics. *Journal of Medical Ethics*, 29(2):103–108, 2003.

[28] Aravind K Joshi, K Vijay Shanker, and David Weir. The convergence of mildly context-sensitive grammar formalisms. *Technical Reports (CIS)*, page 539, 1990.

[29] Anne Kao and Steve R Poteet. *Natural language processing and text mining*. Springer Science & Business Media, 2007.

[30] Jun'ichi Kazama, Takaki Makino, Yoshihiro Ohta, and Jun'ichi Tsujii. Tuning support vector machines for biomedical named entity recognition. In *Proceedings of the ACL-02 workshop on Natural language processing*

in the biomedical domain-Volume 3, pages 1–8. Association for Computational Linguistics, 2002.

[31] Brian W Kernighan and Rob Pike. *The Unix programming environment*, volume 270. Prentice-Hall Englewood Cliffs, NJ, 1984.

[32] Mohammad Reza Khoie, Tannaz Sattari Tabrizi, Shahram Rahimi, Nina Marhamati, et al. A hospital recommendation system based on patient satisfaction survey. *Applied Sciences*, 7(10):966, 2017.

[33] Stephen Cole Kleene. Representation of events in nerve nets and finite automata. Technical report, RAND PROJECT AIR FORCE SANTA MONICA CA, 1951.

[34] Dan Klein and Christopher D Manning. Accurate unlexicalized parsing. In *Proceedings of the 41st annual meeting of the association for computational linguistics*, 2003.

[35] Simon Lacoste-Julien, Fei Sha, and Michael I Jordan. Disclda: Discriminative learning for dimensionality reduction and classification. In *Advances in neural information processing systems*, pages 897–904, 2009.

[36] Anandi V Law, Max D Ray, Katherine K Knapp, and Jeffrey K Balesh. Unmet needs in the medication use process: perceptions of physicians, pharmacists, and patients. *Journal of the American Pharmacists Association*, 43(3):394–402, 2003.

[37] John R Levine, Tony Mason, and Doug Brown. *Lex & yacc.* " O'Reilly Media, Inc.", 1992.

[38] Chenghua Lin and Yulan He. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 375–384. ACM, 2009.

[39] Haibin Liu, Tom Christiansen, William A Baumgartner, and Karin Verspoor. Biolemmatizer: a lemmatization tool for morphological processing of biomedical text. *Journal of biomedical semantics*, 3(1):3, 2012.

[40] Julie Beth Lovins. Development of a stemming algorithm. *Mech. Translat. & Comp. Linguistics*, 11(1-2):22–31, 1968.

[41] L Màrquez and J Giménez. A general pos tagger generator based on support vector machines. *Journal of Machine Learning Research*, 2004.

[42] Jack McKiLLip, Katie Moirs, and Cristine Cervenka. Asking open-ended consumer questions to aid program planning: Variations in question format and length. *Evaluation and Program planning*, 15(1):1–6, 1992.

[43] D Michie, M Nuñez, V Podgorelec, P Kokol, B Stiglic, and I Rozman. Quinlan, jr c4. 5: Programs for machine learning, 1993.

[44] Jeremy R Millar, Gilbert L Peterson, and Michael J Mendenhall. Document clustering and visualization with latent dirichlet allocation and self-organizing maps. In *FLAIRS Conference*, volume 21, pages 69–74, 2009.

[45] Bhaskar Mitra and Nick Craswell. Query auto-completion for rare prefixes. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1755–1758. ACM, 2015.

[46] Antonio Molina and Ferran Pla. Shallow parsing using specialized hmms. *Journal of Machine Learning Research*, 2(Mar):595–613, 2002.

[47] Naomi M Morris. Respect: its meaning and measurement as an element of patient care. *Journal of Public Health Policy*, 18(2):133–154, 1997.

[48] Thomas Müller, Ryan Cotterell, Alexander Fraser, and Hinrich Schütze. Joint lemmatization and morphological tagging with lemming. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2268–2274, 2015.

[49] Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy W Chapman. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5):544–551, 2011.

[50] Preslav Nakov. Improving english-spanish statistical machine translation: experiments in domain adaptation, sentence paraphrasing, tokenization, and recasing. In *Proceedings of the third workshop on statistical machine translation*, pages 147–150. Association for Computational Linguistics, 2008.

[51] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceed-*

ings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, pages 79–86. Association for Computational Linguistics, 2002.

[52] Talcott Parsons. Illness and the role of the physician: A sociological perspective. *American Journal of orthopsychiatry*, 21(3):452–460, 1951.

[53] Adwait Ratnaparkhi. A maximum entropy model for part-of-speech tagging. In *Conference on Empirical Methods in Natural Language Processing*, 1996.

[54] Erik Riiskjær, Jette Ammentorp, and Poul-Erik Kofoed. The value of open-ended questions in surveys on patient experience: number of comments and perceived usefulness from a hospital perspective. *International Journal for Quality in Health Care*, 24(5):509–516, 2012.

[55] Haya R Rubin. Can patients evaluate the quality of hospital care? *Medical care review*, 47(3):267–326, 1990.

[56] Sarah E Schwarm and Mari Ostendorf. Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 523–530. Association for Computational Linguistics, 2005.

[57] Hong Shen and Anoop Sarkar. Voting between multiple data representations for text chunking. In *Conference of the Canadian Society for Computational Studies of Intelligence*, pages 389–400. Springer, 2005.

[58] Madhvi Soni and Jitendra Singh Thakur. A systematic review of automated grammar checking in english language. *arXiv preprint arXiv:1804.00540*, 2018.

[59] R William Soukoreff and I Scott MacKenzie. Measuring errors in text entry tasks: an application of the levenshtein string distance statistic. In *CHI'01 extended abstracts on Human factors in computing systems*, pages 319–320. ACM, 2001.

[60] Ching Y Suen. N-gram statistics for natural language understanding and text processing. *IEEE transactions on pattern analysis and machine intelligence*, (2):164–172, 1979.

[61] Charles Sutton, Andrew McCallum, et al. An introduction to conditional random fields. *Foundations and Trends® in Machine Learning*, 4(4):267–373, 2012.

[62] Yee W Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Sharing clusters among related groups: Hierarchical dirichlet processes. In *Advances in neural information processing systems*, pages 1385–1392, 2005.

[63] Seshadri Tirunillai and Gerard J Tellis. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4):463–479, 2014.

[64] Linlin Wang, Zhu Cao, Yu Xia, and Gerard de Melo. Morphological

segmentation with window lstm neural networks. In *AAAI*, pages 2842–2848, 2016.

[65] Xuerui Wang, Andrew McCallum, and Xing Wei. Topical n-grams: Phrase and topic discovery, with an application to information retrieval. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 697–702. IEEE, 2007.

[66] Yi Wang, Hongjie Bai, Matt Stanton, Wen-Yen Chen, and Edward Y Chang. Plda: Parallel latent dirichlet allocation for large-scale applications. In *International Conference on Algorithmic Applications in Management*, pages 301–314. Springer, 2009.

Index