

# Retrieving Attitudes: Sentiment Analysis from Clinical Narratives

Yihan Deng  
ICCAS  
University of Leipzig  
Simmelweisstr. 14  
Leipzig, Germany

Matthaeus Stoehr  
ENT Clinic  
University Hospital Leipzig  
Liebigstr. 10  
Leipzig, Germany

Kerstin Denecke  
ICCAS  
University of Leipzig  
Simmelweisstr. 14  
Leipzig, Germany

{name.surname}@iccas.de

## ABSTRACT

Physicians and nurses express their judgments and observations towards a patient's health status in clinical narratives. Thus, their judgments are explicitly or implicitly included in patient records. To get impressions on the current health situation of a patient or on changes in the status, analysis and retrieval of this subjective content is crucial. In this paper, we approach this question as sentiment analysis problem and analyze the feasibility of assessing these judgments in clinical text by means of general sentiment analysis methods. Specifically, the word usage in clinical narratives and in a general text corpus is compared. The linguistic characteristics of judgments in clinical narratives are collected. Besides, the requirements for sentiment analysis and retrieval from clinical narratives are derived.

## Categories and Subject Descriptors

H.3 [INFORMATION STORAGE AND RETRIEVAL]:  
Content Analysis and Indexing

## Keywords

Clinical text mining, Sentiment analysis

## 1. INTRODUCTION

Sentiment analysis deals with determining the sentiment with respect to a specific topic expressed in natural language text. So far, the development of sentiment analysis methods concentrated on processing very opinionated, subjective texts such as customer reviews [3, 4]. Clearly, sentiment in clinical documents differs from sentiment in user-generated content or other text types. With the term *sentiment* we refer to information on the health status, or on the outcome of a medical treatment or change / seriousness of a symptom (e.g. *serious pain*) or the certainty of an observation. The work presented in this paper intends to get a more complete

view on the facets of *sentiment* in clinical texts. With the development of the principles of evidence-based medicine [6] and digital patient modeling [1], the observations and judgments expressed in clinical narratives will play a crucial role for the clinical decision process.

Consider the following scenario: During the daily ward round, a physician is making observations with respect to the health status of a patient (e.g. symptoms improved). The patient describes his personal experiences on the symptoms such as the degree of pain. All this information reflects the individual health status and is documented in clinical notes. Retrieving, analyzing and aggregating this information over time can support the treatment decisions and allows a physician to quickly get an overview on the health status. Another application example is retrieving attitudes from clinical documents which can support assessing the outcome of treatments. In this way, labor-intensive user studies for treatment or medication evaluation can be facilitated.

For processing clinical narratives in the last years, effective algorithms in particular for named entity recognition and relation extraction [2] have been developed. Based on recognized entities and relations between entities, sentiments expressed in medical narratives can now be analyzed to offer an upper-level text understanding. Further, a corresponding retrieval of judgments or sentiments can be realized. However, sentiments, opinions and intentions expressed in clinical narratives have not been well exploited yet. In this paper, we start analyzing the sentiment expressions used in clinical texts through a linguistic comparison with a non-medical, subjective text corpus.

Conventional methods for sentiment analysis have been developed for processing subjective on-line documents such as weblogs and forums. In this paper, our goal is to analyze the applicability of such methods for sentiment analysis in clinical narratives. We will identify necessary extensions of existing methods and come up with the requirement of sentiment in clinical narratives. To this end, we will first compare two types of medical narratives (radiology report and nurse letter) with a weblog data set. The lexical and linguistic differences will be presented. Afterwards, we will apply a general subjectivity lexicon to medical narratives using dictionary-based methods. Sources of errors of this simple sentiment recognition approach will be discussed. The following research questions will be addressed:

1. In comparison with user generation content, which lexical characteristics do clinical narratives have?

2. What characterizes sentiments in clinical narratives?
3. Can existing methods for sentiment analysis be applied? Which adaptations are necessary?

## 2. SENTIMENT ANALYSIS IN THE MEDICAL DOMAIN

To our best knowledge, few work considered sentiment analysis in medical texts: Xia et al. [9] have indicated that sentiments are topics-related. Their approach to sentiment analysis starts with a standard topic classifier based on topic labels. In the second step, special classifiers are initialized to detect the polarity for each topic. The multi-step classification method has earned a nearly 10% improvement of F1 measure in comparison with the single-step approach. Niu et al. consider sentiment analysis in biomedical literature [5]. They exploit a supervised method to classify the polarity at sentence level. The linguistic features such as uni-grams, bi-grams and negations are employed. The medical terms are merely replaced by their semantic category. The category information and context information are derived from the Unified Medical Language System (UMLS<sup>1</sup>). The combination of linguistic features and domain-specific knowledge have improved the accuracy of the algorithm. In summary, existing methods for sentiment analysis in the medical domain focus on processing biomedical literature and patient-generated text. The clinical text which is used to record the activities and judgments of health care workers has not yet been analyzed. Moreover, the existing approaches and definitions of sentiment in the medical domain are derived from general sentiment analysis for Web 2.0 media. Clinical context and medical knowledge have not been used thoroughly besides some category meta data derived from the UMLS [7, 5]. We expect that due to different expressions and the more objective way of writing in the clinical narratives, the conventional sentiment analysis methods need to be adapted to cope with the clinical context. We will concentrate on that particular text material.

## 3. METHODOLOGY

### 3.1 Text Material

In order to analyze the differences between the language in clinical narratives and general texts from the Internet, 200 nurse letters and 200 radiology reports from “MIMIC II Database<sup>2</sup>” have been chosen as corpus. These documents form the domain-specific data source in our assessment. For comparison reasons, we additionally consider 200 technical interviews downloaded from the website Slashdot<sup>3</sup>. We have chosen that particular dataset since it belongs to the category of user-generated, subjective content. Given the technical topics, we however expect a certain similarity, mainly an objectivity as it occurs in clinical narratives.

**Nurse Letter:** A nurse letter is part of a patient record, and is written by nurses on duty. Its content reflects the situation of the patient and the feedback to the ongoing treatment. It is written in a relatively subjective manner. Acronyms and typos appear very often in nurse letters.

<sup>1</sup><http://www.nlm.nih.gov/research/umls/>, accessed: 20.04.2014

<sup>2</sup><http://www.physionet.org/>, accessed 20.04.2014

<sup>3</sup><http://slashdot.org>, accessed 20.04.2014

**Radiology Report:** A radiological report is mainly used to inform the treating physicians about the findings in a radiological examination. It starts usually with a medical history, which is followed by a description of the region of interest and questions for the examinations. The texts contain many judgments and observations as observed in the examination.

**Slashdot Interviews:** Slashdot is a technology-related weblog, which covers different technical topics. The users express their opinions on certain topics. We chose the technical interviews as benchmark instead of movie or product review, since technical interviews contain also a relatively large amount of terminologies.

### 3.2 Linguistic and Sentiment Analysis of the Data Sets

Apparently, the three text sources are different in terms of terminology usage and content. The interview corpus is typical user generated content. We expect that the corpus will contain a relatively large amount of sentiment terms and subjective expressions, while the clinical narratives are written in a more objective way. Less opinionated terms and rather more clinical terminology are expected. However, the question is whether the terminology and word usage is really distributed as expected. To what extent do the corpora differ with respect to linguistic characteristics? Recalling our initial research questions, we need to answer whether existing sentiment lexicons can provide the basis for analyzing judgments and sentiments in clinical narratives. In order to address these questions, an extraction pipeline has been built to obtain part of speeches and sentiment terms from the texts and to determine their occurrence frequency. The Penn Tree POS-tagger<sup>4</sup> and the SL sentiment lexicon [8] (contains 8,221 single-term subjective expressions) have been exploited for this purpose. The punctuation, numbers and stop words were also extracted and their proportions were calculated.

After analyzing the linguistic composition of the data sets, we want to study the applicability of a dictionary-based sentiment analysis approach on clinical narratives. Potential limitations of the approach when applied to medical narratives will be identified. For this purpose, we have created an experiment pipeline in KNIME<sup>5</sup>. Two dictionary taggers were applied to recognize positive and negative terms in the text respectively. A voting algorithm is applied to calculate the polarity for each document. It is based on the number of positive and negative occurrences and handles negations. Although it is only a simple approach, it is a direct method to evaluate the compatibility between the subjectivity lexicon and clinical narratives. The SL sentiment lexicon from Wilson et al. [8] is used by the dictionary tagger. It comprises a large amount of adjectives, adverbs, but also nouns and verbs expressing sentiments. For evaluation purposes, the three corpora were annotated with an overall document polarity at document level by one physician from our university hospital.

## 4. RESULTS AND DISCUSSION

### 4.1 Results of the Linguistic Analysis

<sup>4</sup><http://www.cis.upenn.edu/treebank/>, accessed 20.04.2014

<sup>5</sup><http://www.knime.org/>, accessed 20.4.2014

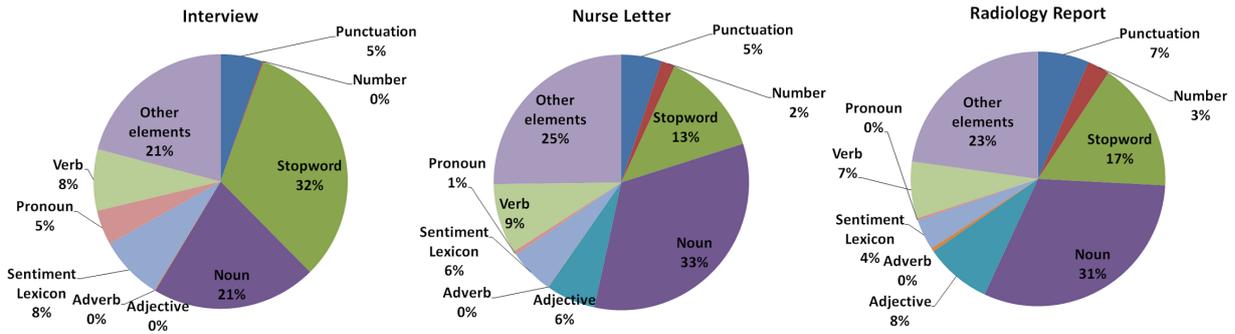


Figure 1: Result of the Linguistic Analysis

In Figure 1, the proportions of punctuations, numbers, stop words, nouns, pronouns, adjectives, adverbs as well as the sentiment terms are illustrated. Part of speeches of terms that matched with the sentiment lexicon have not been considered. The result has partially confirmed our expectation.

*Sentiment Terms:* According to the results, the normal interview corpus contains the highest proportion of sentiment terms with 8%, while the radiology reports contain 4% and nurse letters 6% sentiment terms. These results have approved our observation that nurse letters are written more subjectively in comparison to radiology reports, but they are still more objective than the interviews. The sentiments expressed in nurse letters are normally implicit and appear with the description of patient’s health status, or the social records for the visitors of the patients. Opinionated terms and expressions such as suspicion, negation, approval or recommendations can be found in radiology reports mainly in the conclusion section or impression part at the end of the whole report.

*Number:* Numbers are one of the most important elements in clinical reports, where they are mainly used to represent the dose of medications, the size of a tumor or the frequency of a treatment, etc. In our clinical data sets, numbers comprise between 2% and 3% of the words or characters. In contrast, in the interviews almost no numbers occur, since the discussions in weblogs are more likely to use simple, colloquial vocabulary to present the personal attitudes and preferences.

*Stop Word:* The nurse letters and radiology reports contain 13% and 17% stop words respectively. In contrast, the percentage of stop words in the interview corpus is with 32% significantly higher, which shows that the clinical documents are clearly written in a concise way, focusing on facts.

*Nouns and Pronouns:* What noteworthy is, the percentage of nouns in radiology reports (31%) and nurse letter (33%) is clearly higher than the percentage of nouns in interviews (21%), while the percentage of pronouns in the interviews (4%) is notably higher than in the radiology report (0%) and nurse letters (1%). The reason is that in medical facts are described in clinical narratives using nouns from medical terminologies (e.g. names of diseases, symptoms, medications). In contrast, the interviews contain more subjective terms and use a large amount of first person expressions to express the ideas and opinions of individuals.

*Adjective and Adverb:* Another interesting finding is that the clinical narratives contain a substantial amount of ad-

Types	Accu(overall)	F1(Bad)	F1(Neutral)	F1(Good)
Interviews	0.696	0.754	0.367	0.735
Nurse Letter	0.420	0.437	0.216	0.503
Radiology Report	0.446	0.297	0.080	0.559

Table 1: Sentiment Analysis Results: Accuracy and F1 measure for three text types

jectives (6-8% of the terms) that are not included in the SL sentiment lexicon. In contrast, all adjectives in the interview corpus matched with the sentiment lexicon. The additional adjectives in clinical narratives are mainly related to body locations, such as “left” side, “right” side, “vertical”, “dorsal”, “cervical”. They express neither emotion nor attitude but anatomical concepts and relative locations in the body. In summary, the nurse letters show a relatively higher linguistic similarity to technical interviews than radiology reports. They are to a certain extent more subjectively written than radiology reports. The large amount of the medical terms (noun, adjective) describe the status of a patient. They reflect the attitudes of physicians. Thus, the implicit clinical events may influence the polarity outcome of a clinical report as well. Consequently, the implicit clinical events and evidences are expected to be relevant to understand and interpret the status of the patient.

## 4.2 Results of the Sentiment Analysis

The automatically retrieved polarity for the texts were compared to the manual annotation done by clinical experts. The overall accuracy and F1 measure for the three text types is shown in Table 4.1: Accuracy is the proportion of true results in the population. The sentiment analysis of interviews leads to an acceptable accuracy of 69.6%. The results for nurse letters and radiology reports have merely achieved the accuracies of 42% and 44% respectively. This shows that existing methods need to be adapted when processing these texts and that **sentiment** is different. Furthermore, the F1 measure for positive texts (F1 good) is significantly higher for the clinical texts than for negative (F1 bad) texts. A manual assessment showed that the positive sentiments or outcomes are described in an explicit way, e.g., by phrases such as the “patient slept well, the treatment has a satisfactory result” or “the tube has been placed successfully”. For negative clinical events, the nurse and physician were more

likely to express the status of patient in a careful and cautious manner, e.g. by phrase such as “some situation cannot be excluded or need further pathological investigation”. The radiology reports are more likely to exclude or confirm the occurrences of certain clinical events rather than to give a final diagnosis. In addition, the recognition of neutral situations is difficult, since the judgment of neutral outcome depends on the recognition of positive and negative terms. However, neutral clinical outcomes in real world are probably not objectively expressed. Some surgical result may only show moderate effect, but it may turn out to be an insignificant outcome in nurse letters or might even produce some negative feedbacks. During the annotation, our physician tended to give more positive and negative judgments to the reports rather than neutral ones, since the determination of “neutral” needs more context and reference, which is not that easy to obtain without knowing the complete patient history.

## 5. CONCLUSION AND FUTURE WORK

In this paper, we have studied the linguistic characteristics of clinical narratives compared to a web data set and analyzed the feasibility of a simple sentiment analysis approach on clinical narratives. The results provide important insights to understand **sentiment** in clinical narratives and to continue with developing corresponding analysis methods. The initial three research questions raised in Section 1 can be answered.

1. The linguistic analysis showed that clinical narratives contain a moderate amount of sentiment terms. In contrast to the web data set, more numbers, medical terms (nouns), location-related adjectives are exploited and less stop words, and less pronouns are included. This composition and word usage reflects the objectivity and preciseness of the clinical writing style.
2. By analyzing the clinical documents, we learned more about the nature of sentiment in clinical narratives. Sentiment can concern the general health status of a patient, the outcome of a treatment or of a specific medical condition or can concern uncertainty of an observation. **Good, bad** or **positive** and **negative** is manifested in status changes, e.g. an improvement or worsening of a certain medical or physical condition or the success or failure of a treatment. Sentiment can be seen as health status of a patient: The patient’s health status can be **good, bad** or **normal** at some point in time, expressed either implicitly or explicitly. By analyzing that health status over time, improvements or worsening in the status can be recognized. An implicit description of a health status concerns the mentioning of critical symptoms (e.g. serious pain, extreme weight loss, high blood pressure). A explicit description of the health status is reflected through phrases such as “the patient recovered well” or “normal”. Sentiment in clinical texts can be the outcome of a treatment or the impact of a specific medical condition, i.e. whether the condition improved or worsens which allows to draw conclusions on the effect or outcome of a treatment (positive/negative outcome). The phrase “blood sugar decreased” could express a positive or negative change depending on the previous state. A decrease of blood pressure can be good when it was too high before. This

also shows that for interpreting the detected sentiment, the context need to be considered. Further, sentiment can be seen as presence, change in or certainty of a medical condition. I.e. a medical condition can exist, improve, worsen, be certain or uncertain. The treatment outcome can be positive, negative (e.g. surgery was successful or failed), neutral or a treatment can have no outcome.

3. A simple method for sentiment analysis is not well suited to analyze sentiment in clinical narratives. Sentiment in clinical texts differs significantly from sentiment in general texts. In particular, implicit sentiments need to be detected. An adapted annotation scheme should be defined with the help of physicians. New features for sentiment analysis need to be collected for gathering these subjective sentiments.

In the short term, we will develop a sentiment lexicon specific for the medical domain. It will define a scheme for analyzing and retrieving implicit sentiments and attitudes expressed in clinical texts. The kind of influence and degree of influence of a symptom to the health status will be defined. This lexicon or ontology will be exploited for developing a more comprehensive sentiment analysis algorithm.

## 6. REFERENCES

- [1] K. Denecke. Model-based Decision Support: Requirements and Future for its Application in Surgery. *Biomedical Engineering*, 58(1), 2013.
- [2] C. Friedman, T. C. Rindfleisch, and M. Corn. Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the national library of medicine. *Journal of Biomedical Informatics*, 46(5):765–773, 2013.
- [3] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proc. Tenth ACM SIGKDD*, KDD ’04, pages 168–177, New York, NY, USA, 2004. ACM.
- [4] B. Liu. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on HLT. Morgan & Claypool Publishers, 2012.
- [5] Y. Niu, X. Zhu, J. Li, and G. Hirst. Analysis of polarity information in medical text. In *In: Proc of the AMIA 2005 Annual Symposium*, pages 570–574, 2005.
- [6] D. L. Sackett, W. M. Rosenberg, J. Gray, R. B. Haynes, and W. S. Richardson. Evidence based medicine: What it is and what it isn’t. *BMJ*, 312(7023):71–72, 1996.
- [7] A. Sarker, D. Molla, and C. Paris. Outcome polarity identification of medical papers. In *Proceedings of the Australasian Language Technology Association Workshop 2011*, pages 105–114, Canberra, Australia, December 2011.
- [8] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc of HLT ’05*, HLT ’05, pages 347–354, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.
- [9] L. Xia, A. L. Gentile, J. Munro, and J. Iria. Improving patient opinion mining through multi-step classification. In *TSD*, pages 70–76, 2009.