Artificial Intelligence in Medicine 64 (2015) 17–27

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Position Paper

Sentiment analysis in medical settings: New opportunities and challenges

Kerstin Denecke*, Yihan Deng

Innovation Center Computer Assisted Surgery, University of Leipzig, Semmelweisstr. 14, 04103 Leipzig, Germany

ARTICLE INFO

Article history:
Received 2 September 2014
Received in revised form 18 March 2015
Accepted 25 March 2015

Keywords:
Sentiment analysis
Clinical text mining
Medical language processing
Health status analysis

ABSTRACT

Objective: Clinical documents reflect a patient's health status in terms of observations and contain objective information such as descriptions of examination results, diagnoses and interventions. To evaluate this information properly, assessing positive or negative clinical outcomes or judging the impact of a medical condition on patient's well being are essential. Although methods of sentiment analysis have been developed to address these tasks, they have not yet found broad application in the medical domain.
Methods and material: In this work, we characterize the facets of sentiment in the medical sphere and identify potential use cases. Through a literature review, we summarize the state of the art in healthcare settings. To determine the linguistic peculiarities of sentiment in medical texts and to collect open research questions of sentiment analysis in medicine, we perform a quantitative assessment with respect to word usage and sentiment distribution of a dataset of clinical narratives and medical social media derived from six different sources.
Results: Word usage in clinical narratives differs from that in medical social media: Nouns predominate. Even though adjectives are also frequently used, they mainly describe body locations. Between 12% and 15% of sentiment terms are determined in medical social media datasets when applying existing sentiment lexicons. In contrast, in clinical narratives only between 5% and 11% opinionated terms were identified. This proves the less subjective use of language in clinical narratives, requiring adaptations to existing methods for sentiment analysis.
Conclusions: Medical sentiment concerns the patient's health status, medical conditions and treatment. Its analysis and extraction from texts has multiple applications, even for clinical narratives that remained so far unconsidered. Given the varying usage and meanings of terms, sentiment analysis from medical documents requires a domain-specific sentiment source and complementary context-dependent features to be able to correctly interpret the implicit sentiment.

1. Introduction

In the event of complex medical conditions and associated treatments, patients are examined by different types of specialists. Because they are specialized in a particular medical field, e.g. radiology, surgery, paediatrics etc. this leads to a restricted view of each patient's status and their medical conditions. In their documentations, physicians describe their personal views and observations. This might be a judgement or evaluation, an affective state, or be intended to provoke some emotion in the reader. Sentiment analysis methods aim to determine the attitude expressed with respect to some theme or the overall contextual polarity of a document. Originating in the field of web mining, the development of sentiment analysis methods often concentrates on processing very subjective texts such as customer reviews [1,2]. Limited work has shown that less subjectively written texts can be analysed using sentiment analysis methods, enabling better insights into semantics [3].

In this work, we describe facets and potentials of sentiment analysis in the context of medicine and healthcare for several reasons. A treatment process often involves various persons, including physicians of different specialities, nurses, therapists etc. Since the personal observations and attitudes of a physician influence clinical decision-making, it is crucial to identify them in medical records so that a complete view of a patient's health status can be achieved and presented to other treating healthcare professionals. While examination results are often reported in a structured manner, observations or experiences are communicated in an unstructured way in finding reports or other clinical documents. Accordingly, extracting opinions and intentions from medical narratives can be

* Corresponding author. Tel.: +49 341 97 12002; fax: +49 341 97 12009.
E-mail address: kerstin.denecke@iccas.de (K. Denecke).

http://dx.doi.org/10.1016/j.artmed.2015.03.006
0933-3657/

Table 1
Entities and events in the medical domain and possible sentiment characteristics.

Entity	Possible sentiment values
Health status	Improve, worsen
Medical condition	Present, improve, worsen
Diagnosis	Certain, uncertain, preliminary
Effect of a medical event	Critical, non-critical
Medical procedure	Positive or negative outcome, successful or unsuccessful
Medication	Helpful, useless, serious adverse events

important for assessing clinical data, monitoring a patient's health status or providing automated decision support for physicians. Even though clinical documents are often written in an objective manner, medical conditions impact a patient's life. Sentiment analysis can help determine this impact from written documents.

Natural language processing of clinical narratives has sparked increasing attentions in recent years, resulting in effective algorithms for named entity recognition and relation extraction methods [4]. Based on recognized entities and relations among them, the analysis of opinion and sentiment in clinical narratives can offer a higher-level text understanding [4]. However, the sentiments expressed in clinical narratives have not been well analysed and exploited yet. This paper examines this newly emerging research topic, identifies potential use cases, and summarizes the main open research questions by:

- Determining the current state of the art in sentiment analysis in healthcare settings
- Describing uses and potential of sentiment analysis in medicine
- Characterizing facets of sentiment in the domain of medicine and drawing conclusions for technical developments in this field

In the following section, we describe the facets of sentiment in the medical context. Section 3 summarizes existing work on sentiment analysis. To characterize sentiment, we analysed and compared the linguistic peculiarities of clinical and medical social media texts with respect to subjectivity and opinions (Section 4). Possible use cases were collected in discussions with physicians and are described in Section 5. Additionally, research challenges for the future development of medical sentiment analysis methods are outlined. The paper finishes with conclusions in Section 6.

2. The notions of sentiment in medicine

Textual information can be broadly categorized into facts and opinions [2]. Facts are objective expressions about entities or events. Opinions are usually subjective expressions describing people's attitudes, sentiments or feelings about entities. However, the concept of sentiment or opinion is quite broad, encompassing subjectivity, polarity, emotion or even comparison. While it is straightforward for a person to like or dislike a movie or product, sentiment in the context of medicine is difficult to capture in a few words. Facets of sentiment in health-related texts may concern (see Table 1):

- A **change** in health status (e.g a patient can suddenly feel better or worse)
- **Critical events, unexpected situations or specific medical conditions** that impact the patient's life (e.g. *tumour is malignant* as such is a fact, but this medical condition is negative for the patient since it might lead to health problems or death)
- The **outcome or effectiveness of a treatment** (e.g. surgery may be successfully completed)

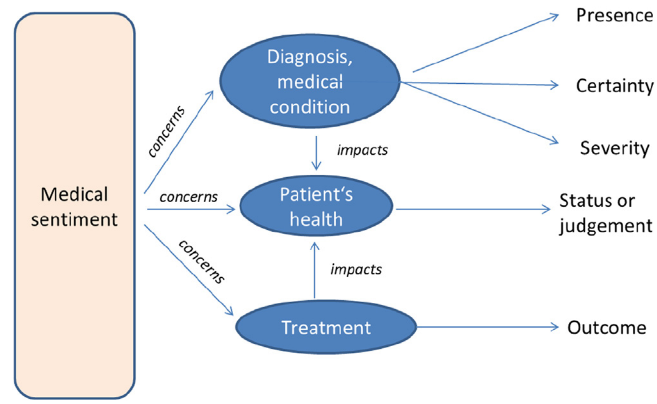

Fig. 1. Facets of sentiment in medical contexts. Sentiment can concern patient's health, a medical condition or treatment. For each aspect, sentiment can occur in different forms.

- **Experiences or opinions of a treatment or a sort of drug** (e.g. a patient or a physician can describe serious adverse events after drug consumption)
- The **certainty of a diagnosis** (e.g. a physician may be certain of some diagnosis)

The examples show that *good* and *bad* or *positive* and *negative* in the context of medicine concern the health status, a medical condition or a form of treatment. It is manifested in improvements or deteriorations of certain medical or physical conditions or in the success or failure of a treatment (see Fig. 1). Below these facets are described in more detail.

Sentiment can be seen as a **reflection of the health status** of a patient which can be *good*, *bad* or *normal* at some point in time. Thus, the health status impacts the quality of life of a patient (e.g. a *severe pain* affects the life of a patient much more than a *slight pain*). By analysing health status over time, improvements or deteriorations in the status can be recognized. In clinical narratives, the health status is expressed either implicitly or explicitly. Implicit descriptions of the health status concern the mentioning of symptoms (e.g. *severe pain, extreme weight loss, high blood pressure*). They require additional content information for correct interpretation. An explicit description of the health status is reflected in phrases such as *the patient recovered well* or *normal*.

A medical condition can exist, improve or worsen. Thus, sentiment can also be considered as the **presence or change of a medical condition**. Sentiment in this context can be implicitly considered as the severity of a disease, which again impacts on life circumstances. A medical condition can have different weights: a chief complaint might affect the health status much more than another symptom. Furthermore, the **certainty of a diagnosis** can be seen as the opinion of a physician. For example, a diagnosis may be a suspicion or it may be assured. This opinion on the certainty of a diagnosis impacts the treatment decision: if a diagnosis is certain, a treatment decision can be made; otherwise additional examinations are necessary. Another interesting facet of sentiment concerns the **judgement of medical conditions**, in particular with respect to their severity. For instance, events such as a *bleeding* can be positive or negative, critical or less critical. The phrase *blood pressure decreased* could express a positive or negative change depending on the previous state of blood pressure. A decrease of blood pressure can be good if it was too high before. This also shows that sentiment in clinical narratives cannot always be manifested in single terms or phrases, and that the context is important.

Additional sentiment aspects concern treatment. It can be complex or less complex, urgent or less urgent. The **outcome of a treatment** may be *positive, negative* (e.g. surgery was successful or failed), *neutral* or a treatment can have *no outcome*. The outcome can often only be derived from the described effects of a treatment on a medical condition. For example, a statement that a medical condition has improved allows the conclusion that the treatment had a positive outcome. **Observations and opinions** of treatments or medications expressed in clinical narratives or in social media documents provide another facet of sentiment in the context of medicine.

In conclusion, the concept of medical sentiment is very complex and has multiple facets making it very interesting, but also challenging for automatic analysis. Through a linguistic analysis in Section 4, we will collect the challenges to be considered.

3. Approaches to sentiment analysis

In the context of web science and web mining, methods for determining and analysing subjectivity as well as opinions and emotions expressed in written format have been developed. The field is known as opinion mining, or sentiment analysis. In medicine and health, sentiment analysis is attracting growing interest. In this section, we first provide a brief overview of sentiment analysis in general. Then we describe and classify more deeply the existing approaches to sentiment analysis in the context of medicine.

3.1. General overview

Research in sentiment analysis started in around 2004 [1] with the mining of opinions in customer reviews and online news. The initial task considered was to distinguish *positive* from *negative* reviews (referred to as **polarity analysis**) or *subjective* from *objective* parts of a text (referred to as **subjectivity analysis**). Later on, additional tasks were introduced: **emotion analysis** determines the emotional category of texts (e.g. *anger, disgust*) [5,6] while **intensity analysis** focuses on identifying different levels of polarity or emotion (e.g. *very positive, very sad*).

Existing sentiment analysis methods developed for processing unstructured text in web media treat the task as a classification problem: a classifier is trained to detect the polarity at sentence or document level. For example, Pang et al. presented a series of supervised methods such as naive Bayes, support vector machines (SVM) and maximum entropy classification [7,8]. An unsupervised mechanism has been developed by Turney, who proposed a solution to recognize the semantic orientation of texts [9].

These approaches are based on feature sets including semantic or lexical features. Often, the underlying assumption is that sentiment is explicitly mentioned in the text, and is thus manifested in opinionated words. Normally, adjectives, adverbs and specific nouns express sentiments in free texts. Given this, conventional sentiment analysis is based on the detection and analysis of such opinionated terms (semantic features). These terms are usually categorized into the categories *positive, negative* and *neutral* and made available in sentiment lexicons (see Section 3.2). The latter are then exploited by sentiment analysis algorithms to identify the opinionated terms and their polarity in a text through lexicon-lookup.

Apart from a lexicon-lookup to identify opinionated terms, additional features can be extracted from texts for exploitation in sentiment analysis methods. They include lexical features such as unigrams, bigrams and parts of speech.

Sentiment analysis can be considered at different levels: at the level of a word, aspect, sentence and document. Pang et al. [7] considered the problem at document level. However, there is a need for fine-grained approaches to sentiment analysis. Aspect-based analysis aims to identify the aspects of entities and assigns a sentiment to each aspect (e.g. *The CD is good but it was too expensive* describes two aspects, the content of the CD and its price) [10].

Studies have shown that sentiment analysis is often domain-dependent [11] since the polarity of single terms can differ depending on the context they are used in. For that reason, feature models on which a machine learning classifier is based, need to be trained on domain-specific datasets. Lexicons need to be adapted to the domain-specific interpretations of words. An approach to domain-independent sentiment analysis was presented by Montejo-Raez et al. [12]. They represented each term in a tweet as a vector of weighted WordNet synsets that are semantically close to the term. The weights were used in SentiWordNet [13] to estimate the polarity.

Limited work is available on sentiment analysis from fairly objective texts. Balahur et al. applied sentiment analysis methods to news articles to separate the good and bad news content from the good and bad sentiment expressed on a target and to mark explicit opinions [3]. For this particular text type of news articles, opinion is "conveyed through facts that are interpretable by the emotion they convey" [3]. This is to a certain extent comparable to sentiment analysis in healthcare: sentiment can be conveyed through diseases, treatments or medical conditions and they impact a patient's life quality and health status.

3.2. Sentiment lexicons

Many approaches to sentiment analysis are based on sentiment lexicons. They provide an important basis for recognizing sentiment terms and patterns of sentiment expressions in natural language texts. Existing lexicons include SentiWordNet (SWN) [13], WordNetAffect, General Inquirer[1] and the Subjectivity Lexicon (SL) [14]. They contain words and assigned sentiment scores or classes to single terms. The lexicons can be generated manually or through corpus analysis. Often, they do not consider the peculiarities and differing meanings of terms when used in different domains or do not explicitly provide this information regarding other domains. SentiWordNet is one of the most widely used sentiment lexicons. It assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity. SentiWordNet contains different meanings of terms. However, it is not specified in which domain a term has another meaning (e.g. if the term *right* is just employed directionally or used in the sense of "correct"). WordNetAffect [15] provides to a number of WordNet synsets one or more affective labels representing emotional states. Mohammad and Turney [16] developed EmoLex, an emotion lexicon generated by manual annotation through Amazon's Mechanical Turk service. The Subjectivity Lexicon of Wilson et al. [14] contains 8221 single-term subjective expressions associated with their polarity.

In addition to using such general sentiment resources, domain-specific lexica are going to be developed which consider the domain-dependency of meanings of words. A high-quality multi-dimensional lexicon should include the most representative terms in the domain with their semantic relations and high-level event scheme. The extension of an existing lexicon is the most direct way to cover the domain-specific context. One approach is to increase the lexicon coverage by merging a general lexicon with a domain-specific lexicon.

While conventional methods have mainly focused on determining explicit sentiment expression, the annotation scheme proposed by Deng et al. [17] considered the benefactive and malefactive events with respect to the opinionated entities and created an extension of the well-known subjectivity lexicon (MPQA) [14]. More specifically, four types of the goodFor and badFor events (gfbf) were defined: *destruction, creation, gain or loss, benefit or injury*. The

[1] http://www.wjh.harvard.edu/~inquirer/ (accessed 08.02.15).

events are represented as a triple of text spans ⟨*agent*, *gfbf*, *object*⟩, which indicates that some subject (noun) such as a person or an organization had some good or bad effect on the object (noun). The implicit opinions expressed by the writer are presented in the scheme.

Goeuriot et al. [18] merged terms from the general lexicons SentiWordNet and Subjectivity Lexicon [14]. Afterwards, opinionated terms from drug reviews were extracted to extend the merged lexicons. During extension, the difference of polarity between the general domain and the medical domain was analysed. They found out that some words considered generally as *neutral* are opinionated in medical texts. Finally, an evaluation with a simple voting algorithm showed that better results can be achieved with the merged lexicon than with well-known general sentiment lexicons.

Ohana et al. [19] evaluated the performance of sentiment detection using four lexicons: General Inquirer [20], Subjectivity Lexicon [14], SentiWordNet and Moby.[2] The lexicons were applied to texts from six general domain datsets including book reviews, hotel feedback and discussions about music and films. The results showed that the accuracy of the single lexicons depends on the domain. More specifically, SentiWordNet scored the highest accuracy (65–71%) in four out of six domains, while the Subjectivity Lexicon has provided the best performance (63–65%) in the two other domains. One reason for this is the large vocabulary coverage of SentiWordNet.

3.3. Sentiment analysis in a medical context

Similar to general sentiment analysis approaches, existing research on sentiment analysis in the context of medicine can be grouped according to textual source (e.g. medical web content, biomedical literature, clinical note), task (e.g. polarity analysis, outcome classification), method (e.g. rule-based, machine-learning based) and level (e.g. word level, sentence level). The main approaches are described in more detail in the following.

3.3.1. Sentiment analysis from the medical web

Most research on sentiment analysis in the domain of medicine considers web data such as medical blogs or forums for the purpose of mining or studying patient opinions or measuring quality. For example, a method was introduced that separates factual texts from experiential texts [21] to measure content quality and credibility in patient-generated content. Assuming that factual content is better than affective content since more information is given (in contrast to moods and feelings), a system has been developed using subjectivity words and a medical ontology to evaluate the factual content of medical social media.

As in general sentiment analysis, existing approaches to sentiment analysis from medical web data are either machine-learning based [22,23] or rule-based [24]. Most of the work focuses on polarity classification. Xia et al. introduced a multi-step approach to patient opinion classification [23]. Their approach determines the topic and the polarity expressed towards it. An F-measure of around 0.67 was reported. Sokolova et al. tested several classifiers including naive Bayes, decision trees and support vector machines for the sentiment classification of tweets [25]. Texts were represented as bags of words. Two classification tasks were considered: three-class (positive, negative and neutral) and two-class (positive, negative). The best F-measure of 0.69 was achieved with an SVM classifier.

The work by Biyani et al. [26] focused on determining the polarity of sentiments expressed by users in online health communities. More specifically, they performed sentiment classification of user posts in an online cancer support community (cancer survivors network) by exploiting domain-dependent and domain-independent sentiment features as the two complementary views of a post and exploit them for post-classification in a semi-supervised setting employing a co-training algorithm. This work was later extended with features derived from a dynamic sentiment lexicon, whereas the previous work used a general sentiment lexicon to extract features [27].

Smith et al. [28] studied another aspect of sentiment in patient feedback, namely discourse functions such as *expressive* and *persuasive*. A classifier was evaluated based on a patient feedback corpus from NHS Choices.[3] The results illustrate that the multinomial naive Bayes classifier with frequency-based features can achieve the best accuracy (83.53%). Further, the results showed that a classification model trained solely on an *expressive* corpus can be directly applied to the *persuasive* corpus and achieve a comparable performance as the training based on the corpus with the same discourse function. Another interesting application of sentiment analysis was presented by Sharif et al. [29]. Their framework extracts important semantic, sentiment and affect cues for detecting adverse drug events reported by patients in medical blogs. This approach is able to reflect the experiences of people when they discuss adverse drug reactions as well as the severity and emotional impact of their experiences.

Na et al. [24] presented a rule-based linguistic approach for the sentiment classification of drug reviews. They used existing resources for sentiment analysis, namely SentiWordNet and the Subjectivity Lexicon [14], and came up with linguistic rules for classification. Their approach achieved an F-measure of 0.79.

Additional work has tackled the detection and analysis of emotion in medical web documents. Sokolova and Bobicev [30] considered the categories *encouragement* (e.g. hope, happiness), *gratitude* (e.g. thankfulness), *confusion* (e.g. worry, concern, doubt), *facts*, and *facts+encouragement*. They used the affective lexicon WordNetAffect [15] for the emotion analysis of forum entries and achieved with a naive Bayes classifier an F-measure of 0.518. Melzi et al. [31] applied an SVM classifier on a feature set comprising unigrams, bigrams and specific attributes to classify sentences into one of six emotion categories.

3.3.2. Sentiment analysis from biomedical literature

In addition to medical social media data, biomedical literature has been analysed with respect to the outcome of a medical treatment. In this context, sentiment refers to the outcome of a treatment or intervention. Four classes were considered in existing work: *positive, negative, neutral outcome* and *no outcome* [32]. Niu et al. exploited a supervised method to classify the (outcome) polarity at sentence level. Unigrams, bigrams, change phrases and negations and categories were employed as features. According to the results, the algorithm's accuracy was improved by the usage of category information and context information derived from a medical terminology, the unified medical language system (UMLS[4]).

Sarker et al. developed a new feature called the relative average negation count (RANC) to calculate polarity with respect to the number and position of the negations [33]. This count suggests that a larger total number of negations reflects a negative outcome. The experimental corpus was collected from medical research papers, which are related to the practice of evidence-based medicine. An N-gram feature set with RANC exploited by an SVM classifier achieved an accuracy of 74.9%.

[2] http://icon.shef.ac.uk/Moby/ (accessed 09.02.15).

[3] http://nhs.uk (accessed 08.02.15).

[4] http://www.nlm.nih.gov/research/umls/ (accessed 08.02.15).

3.3.3. *Sentiment analysis from other medical texts*

Some researchers concentrated on additional sources of medical texts to apply sentiment analysis or emotion detection methods. A comprehensive shared task of sentiment analysis based on suicide notes was addressed in an i2b2 challenge [34]. The best performance of an F_1 measure of 0.61 was achieved with an SVM classifier and pattern matching. However, only some of the participating teams of the challenge have published the details of their feature engineering and selection of algorithms.

Cambria et al. [35] introduced Sentic PROMs, a concept where emotion analysis methods were integrated into a framework to measure healthcare quality. In a questionnaire, patients answered questions regarding their health status. From the free text entered, emotion terms such as "happy" and "sad" were detected using the semantic resources ConceptNet [36] and WordNet-Affect [15]. The extractions were assigned to one of 24 affective clusters following the concept of *hourglass of emotions* [37]. This concept presents the affective common sense knowledge in terms of a vector, which shows the location in the affective space.

3.4. *Summary of medical opinion mining approaches*

In summary, existing methods for sentiment analysis in the medical domain focus on processing web content or biomedical literature. The clinical narratives which are used to record the activities and observations of physicians as well as patient records have not yet been analysed in this context. In terms of methods, rule-based approaches are presented, but the majority of papers reports on machine-learning methods (SVM [34], naive Bayes [25], regression tree) using features such as parts of speech and uni-, bi- and trigrams. Although general sentiment lexicons are exploited, experiments showed that they are not well suited for capturing the meanings in medical texts. In contrast to "normal" sentiment analysis, additional domain-specific features have been explored in some approaches, mainly UMLS concepts reflecting medical conditions and treatments. The main tasks considered have been polarity classification, but new tasks are emerging including outcome classification, information content classification or emotion analysis. However, the existing work on medical sentiment analysis does not cover all facets of sentiment analysis described in Section 2. In summary, there is still a huge potential for future research. We will outline the main challenges in Section 5. Table 2 summarizes the related work in medical sentiment analysis.

4. Sentiments in medical texts: a quantitative analysis

This section describes the quantitative analysis we performed to study sentiment expressions in medical texts. First, we describe the data material and method. Afterwards, the analysis results are presented.

4.1. *Method and datasets*

The purpose of our analysis is to analyse the language and sentiment expressions in clinical narratives and medical social media. For this purpose, we created a domain-specific corpus comprising clinical documents (nurse letters, radiology reports and discharge summaries). These documents were collected from the MIMIC II Database.[5] Additionally, we collected drug reviews from DrugRatingz.com[6] and medical blogs from WebMD.[7] Both sources provide the medical social media corpus. For comparison reasons, we also considered technical interviews downloaded from the website Slashdot.[8] We chose this particular dataset from the technical domain since it belongs to the category of user-generated, subjective content. Given the technical topics, we expected a certain similarity to clinical narratives. One thousand entries were gathered from each data source to perform the experiment. Each document contained 500 words on average. In the following, the document categories are described in more detail.

Nurse letter: A nurse letter is part of a patient record, which is written by nurses on duty. Its content comprises observations by the nurses while monitoring patients. It therefore contains information on the patient's response to treatment and their health status. It is written in a relatively subjective manner in comparison to other clinical narratives. Acronyms and typing errors appear very often in nurse letters due to the time pressure in the daily work of nurses.

Radiological report: A radiological report is mainly used to inform the treating physician about the findings of a radiological examination. It usually starts with an anamnesis, which is followed by a description of the region of interest and of observations on the inspected items. The text contains judgements and observations made during the examination.

Discharge summary: A discharge summary includes all the important aspects of a patient's hospital stay. It normally starts with the patient information and medical history, which is followed by the diagnosis and applied medical interventions. Prescribed drugs, conclusion of the treatment process as well as suggestions for further treatment constitute the last part of the summary.

Drug reviews: In drug reviews, users can anonymously rate drugs with respect to several categories including effectiveness, side effects, convenience and value. Additionally, users can post and read comments. These comments, dealing with symptoms and side effects, provided the dataset for our analysis while the ratings remained unconsidered. Drugratingz.com is part of a family of websites dedicated to helping consumers in finding the best businesses, destinations, and services by sharing ratings and reviews.

MedBlogs: The blog posts in our dataset (referred to as "MedBlog" in the remainder of this paper) are collected from the WebMD webpage. WebMD provides health information and forums on its website. The blogs are maintained by physicians to facilitate the dissemination of medical knowledge and provide advice. blog postings from that page.

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

Within a **quantitative comparison**, we analysed and compared the word usage of the six text types. The MedBlogs, Slashdot interviews and the drug reviews are typical user-generated content. We expected these three corpora to contain a large amount of sentiment terms and subjective expressions, while the clinical narratives were expected to be written more objectively with fewer opinionated terms and more clinical terminology. However, the question is whether the terminology and word usage are really distributed as expected. To what extent do the corpora differ with respect to linguistic characteristics? Recalling our initial research questions, we also need to answer whether existing sentiment lexicons can provide the basis for analysing judgements and sentiments in clinical narratives.

5 http://www.physionet.org/ (accessed 28.10.14).
6 http://www.drugratingz.com/ (accessed 20.07.14).
7 http://www.webmd.com/ (accessed 05.11.14).
8 http://slashdot.org (accessed 20.07.14).

Table 2
Summary of sentiment analysis work in the domain of medicine.

Textual source	Paper	Task	Method	Level	Resource	Application
Blogs	Denecke [21]	Information content	Logistic regression	Document	SentiWordNet, UMLS	Quality of social media
Blogs	Xia et al. [23]	Polarity	Multinomial naive Bayes classification	Topic	–	Patient opinion mining
Blogs	Sharif et al. [29]	Polarity	K-means clustering	Document	SentiWordNet	Detecting adverse drug events
Tweets	Sokolova et al. [25]	Polarity	Naive Bayes, decision trees, k-nearest neighbor, SVM	Document	–	Mining personal health information
Forums	Biyani et al.[26]	Polarity	Co-training algorithm	Document	Adapted SL, Wikipedia	Analysing emotional effects
Forums	Tanveer et al. [22]	Polarity	Naive Bayes, SVM, logistic regression	Sentence	SL	Categorization
Forums	Sokolova and Bobicev [30]	Emotion	Naive Bayes,k-nearest neighbor	Document	WordNet Affect	Studying sentiments in forums
Forums	Melzi et al. [31]	Emotion	SVM	Document	EmoLex[16]	Patient knowledge retrieval
Drug reviews	Na et al.[24]	Polarity	Rule-based linguistic	Aspect (overall opinion, effectiveness, side effects, condition, cost, dosage)	SentiWordNet, SL, UMLS	Drug opinion mining
Suicide notes	Pestian et al. [34]	Emotion	Supervised learning	Token and sentence	–	Emotion analysis
Questionnaires	Cambria et al. [35]	Emotion	Rule-based and clustering	Document	ConceptNet, WordNet Affect	Measuring healthcare quality
Questionnaires	Smith et al.[28]	Polarity	Naive Bayes, SVM	Document	–	Sentiment classification, determining relevance of discourse function
Biomedical literature	Niu et al. [32]	Outcome	SVM	Document	UMLS	Determining clinical outcome

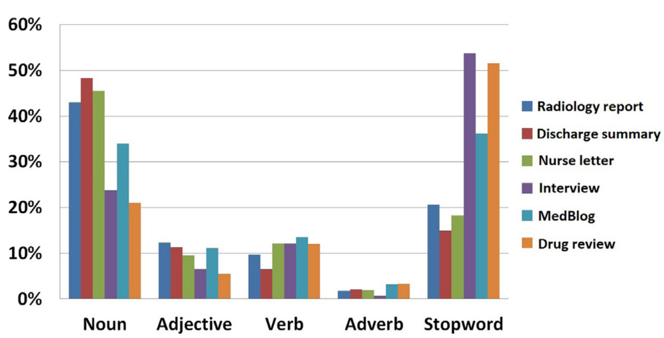

Fig. 2. Quantitative assessment of parts of speech in the six corpora.

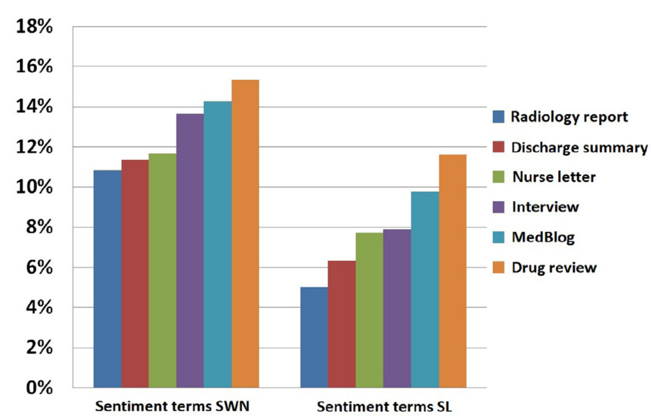

Fig. 3. Quantitative assessment of sentiment terms in the six corpora.

To address these questions, an extraction pipeline was built to quantify parts of speech, to determine their occurrence frequency, and to calculate term matches with sentiment lexicons. The Penn Tree POS-tagger,[9] the Subjectivity Lexicon [14] (containing 8221 single-term subjective expressions) and SentiWordNet (117,659 single terms) were used for this purpose.

4.2. Results of the quantitative assessment

In Fig. 2, for each text source the proportions of nouns, adjectives, verbs, adverbs and stop words are illustrated. Furthermore, Fig. 3 shows the percentage of sentiment terms extracted using dictionary-lookup in the aforementioned sentiment lexicons (SWN and SL). The six text sources are evidently different in terms of terminology usage and content. A more detailed discussion per word class follows.

4.2.1. Nouns

Compared to the other word classes, nouns are the most frequently used parts of speech in all datasets. Strikingly, the percentage of nouns in clinical documents (e.g. radiology reports (42.9%) and discharge summaries (48%)) is clearly higher than in the other text sources (e.g. in interviews (23.7%) or drug reviews (21%)). The reason is that medical facts are described in clinical narratives in a very compact manner using a “nominal” style (e.g. names of diseases, symptoms, medication and treatment are listed). They summarize facts whereas the other document types are rather observational. In drug reviews, the number of nouns is particularly low. This is because mainly impressions and feelings are expressed in these reviews. Interestingly, the number of nouns in the MedBlog dataset is also quite high (34%) compared to the other social media datasets.

[9] http://www.cis.upenn.edu/~treebank/ (accessed 20.04.14).

4.2.2. Adjectives

Another interesting finding is that the clinical narratives and texts from medical blogs contain a substantial amount of adjectives (9–12% of the terms). In contrast, adjectives occur less frequently in the interview corpus and in drug reviews. Through manual inspection, we noted that the adjectives used in clinical narratives and MedBlogs are mainly related to body locations, such as "left" side, "right" side, "vertical", "dorsal" and "cervical". Instead of emotion or attitude, they express anatomical concepts and relative locations in the body.

4.2.3. Adverbs and verbs

With respect to verbs, we can recognize with the exception of discharge summaries, between 10% and 14% of the terms are verbs. In discharge summaries, significantly fewer verbs were identified. This again reflects the fact that discharge summaries are written in a more nominal style listing facts, while the other text types also report actions and behaviours. The number of adverbs is very low for all datasets.

4.2.4. Stop words

Interestingly, the proportion of stop words is very large in the social media datasets with 51% in the drug reviews, 36% in the MedBlogs and as high as 53% in the interview dataset. In contrast, the nurse letters, radiology reports and discharge summary contain 18%, 20% and 14% stop words, respectively. This demonstrates that the clinical documents are clearly written in a much more concise way, focusing on facts and resisting on the use of meaningless words.

4.2.5. Sentiment terms

The analysis of term matches with the two sentiment lexicons (SL, SWN) shows similar trends for both lexicons (see Fig. 3): the drug reviews contain the highest proportion of sentiment terms with 11.6% (SL) and 15.3% (SWN), followed by the MedBlog with 9.7% (SL) and 14.2% (SWN). In contrast, far fewer words in the radiology reports and in the discharge summaries match the SL and SWN. This confirms the more subjective language in social media datasets. Differences are even to be seen between the clinical narratives: nurse letters contain more opinionated terms than the other two clinical datasets. They are written more subjectively than radiology reports, but still more objectively than the social media data. This result confirms our initial hypothesis that nurse letters are more subjectively written than other clinical narratives. After manually reviewing the text material, we noted that the sentiments expressed in nurse letters are normally implicit and appear together with the description of the patient's health status, or the social records for the visitors of patients. Opinionated terms and expressions such as suspicion, negation, approval or recommendations can mainly be found in radiology reports in the conclusion section or impression part at the end of the whole report. In contrast, sentiment terms are distributed throughout the text in the social media datasets.

Another interesting result is that the two lexicons evidently have different coverage. Far more sentiment terms were identified by applying SWN. To investigate the identified sentiment terms, we determined the proportion of adjectives and nouns that were extracted as sentiment terms by applying both lexicons. Second, we identified the top 10 frequent terms per lexicon and corpus (see appendix) to judge their concrete content.

Through these statistics, we learned that matches with SentiWordNet occur frequently with nouns: around 17–47% of the sentiment term matches with SWN are nouns whereas only 7–14% of the matches with SL are nouns. The top frequent terms confirmed again that SWN already covers a certain amount of medical terms, contrasting with the SL result. For instance, in the radiology reports, nouns such as "pneumothorax" and "artery" were identified as sentiment terms when using SWN. Terms frequently extracted using the SL included adjectives such as "significant" and "well". In addition, frequent matches with SWN were adjectives describing anatomical locations such as "bilateral", "upper","subclavian" and "inferior". Analysis using the SL did not recognized these terms. We conclude that SWN matches more terms in the medical documents. However, the result sets also contain many objective terms. This is in contrast to matches recognized using the SL, where only subjective terms are included and therefore extracted. Moreover, the top frequency term statistics also indicated that the polarity of certain adjectives needs to be adapted before application to clinical documents. For example, the terms "patient", "normal", and "patent" have a different polarity in SWN than expected.

Additionally, we studied the proportions of positive and negative terms that are identified by applying the two lexicons (see Table 3). Both lexicons show very different trends:

- Much fewer sentiment terms are detected using SL
- While clinical narratives contain more positive terms when applying the SL, the trend is reversed when using SWN

More specifically, in radiology reports, we identified 61% positive terms (using SL) that were used to describe the findings, while the nurse letters used 58% positive terms to represent the patient's status. This again confirms the hypothesis that medical practitioners are more likely to express negative things in a subtle, implicit manner whereas the vocabulary used in drug reviews and MedBlog is normally incisive and straightforward (meaning that many negative terms are contained). In contrast, a more balanced percentage of positive and negative terms was found in medical narratives when using SentiWordNet. In conclusion, none of the two lexicons seems ideal for analysing sentiment in medical documents. We will discuss this challenge in the next section.

5. The future of sentiment analysis in medicine

In this section, research aspects to be considered in future are summarized that were derived from the analysis presented in the section before. Beyond, we outline possible application areas of sentiment analysis in medicine to indicate the potentials of that research field for the future.

5.1. Research aspects

5.1.1. Domain-specific sentiment source

The application of the two sentiment lexicons to the datasets did indeed show that sentiment terms can be found in clinical narratives. However, we noted that terms such as *right* or *patient* are identified as polarity terms due to their semantic ambiguity. In the medical context, these terms are used objectively. Moreover, the sentiment lexicon SentiWordNet also matched many nouns which are labelled objective. SL originates from social media sources. It contains basic sentiment terms whereas SWN was built on WordNet synsets. Its coverage of terminology has already been proved by its larger recall compared to SL. In order to constitute a suitable lexicon with high precision for the medical domain, more context and domain knowledge should be considered to reduce ambiguity during the matching process. Due to different language use, and the more objective style of writing in the clinical narratives, the conventional sentiment lexicons need to be adapted to cope with these peculiarities of medical narratives. Consider the word "positive". In clinical language, this term is often employed differently from our normal usage. A "positive finding" often has negative

Table 3
Percentage of sentiment terms (positive and negative) extracted from the six sources using SL and SWN.

Sources	Subjectivity lexicon			SentiWordNet		
	Positive	Negative	Total	Positive	Negative	Total
Radiology report	61%	39%	5%	46%	53%	10.8%
Nurse letter	58%	42%	8%	51%	49%	11.6%
Discharge summary	52%	48%	6%	48%	52%	11.3%
Interview	52%	48%	8%	56%	44%	13%
Drug review	48%	52%	12%	46%	54%	15.3%
MedBlog	51%	49%	10%	56%	43%	14%

implications for a patient (e.g. a positive HIV test means that the patient is infected with HIV). Furthermore, the ambiguity and polysemy in the medical sphere are different to the normal domain. Additionally, the implicit aspects of polarity for medical terminology need to be considered (see Section 5.1.3).

The results showed that clinical documents are low in “classical” opinionated terms such as adjectives, and that instead nouns predominate. In order to address this challenge, resources need to be developed that link medical conditions to sentiment or their respective judgements. Such a domain-specific resource could be created based on a fundamental medical ontology, enriched with polarity information and the corresponding inference approaches for concept dependency. Context could also be learned from training material. Our idea is to build a domain-specific lexicon based on concepts of the UMLS, i.e. assigning polarity values to UMLS concepts and extending SentiWordNet with respect to domain-specific polarity in particular for nouns in the medical domain. We have already organized a small annotated dataset.

5.1.2. Level of analysis

Given the multiple notions of sentiment presented in Section 2, it becomes clear that the level of analysis needs to be carefully selected. In general, sentiments can be studied at document, sentence or topic level. For sentiment analysis from clinical narratives, topic- or aspect-level sentiment analysis should be chosen. A clinical document summarizes plenty of information on the health status of a patient; it needs not be restricted to one single topic. Authors of clinical documents express their attitudes and observations with respect to certain body parts or medical conditions of one patient at sentence level. Interpretation of sentiment and judgements in medical sentiment analysis also requires consideration of the context (which means sometimes looking beyond document boundaries). The particular challenge that goes beyond existing research on topic- or aspect-level sentiment analysis, is that a domain-specific knowledge base needs to be incorporated that supports the aggregation of these aspects. For example, problems with a stomach might be described with multiple symptoms: *diarrhoea, abdominal pain, nausea*. Background knowledge is necessary, to realize that these symptoms belong to the same medical condition.

Additionally, the sentiment categories need to be carefully selected depending on the sentiment notion under consideration. The concrete patient status implied by medical conditions cannot simply be judged as positive, negative or neutral. For example, a disease may not be fatal for the patient but may still be very painful. Is this positive or negative? Some other disease may not cause any pain to the patient, yet still be urgent and severe. In order to represent these situations, more detailed aspects should be given instead of conventional polarity (positive or negative). In particular, when we are considering medical conditions, aspects of sentiment include *presence, certainty* and *severity* (see Section 2). New categories of sentiment need to be specified.

5.1.3. Explicit and implicit sentiments

Our analysis confirmed that clinical narratives differ from social media data in terms of word usage. In contrast to sentiments in non-medical social media, where opinion and polarity are manifested in corresponding words (mainly adjectives), in clinical narratives sentiment is often contained implicitly and needs to be inferred for instance from the medical concepts used in documents. Implicit descriptions of the health status mention critical symptoms (e.g. *severe pain, extreme weight loss, high blood pressure*). An explicit description of the health status is reflected in phrases such as *the patient recovered well* or *normal*. In particular, the outcome of a treatment can often only be derived from the described effects of a treatment on a medical condition. For example, a statement that a medical condition improved allows the conclusion to be drawn that the treatment had a positive outcome. This impacts the necessary analysis methods. More knowledge-relevant and context-dependent features need to be chosen to cope with the characteristics of text in the medical domain.

We learned from analysing the clinical narratives, that health status information is described in a very subtle manner by physicians in clinical narratives. Medical language usage is more implicit and conservative to avoid misunderstandings and disputes. Analysis of these subtle sentiments is similar to the analysis of metaphor and sarcasm in texts. Corresponding methods should be analysed and tested.

5.1.4. Summary

In summary, the following research challenges need to be addressed when analysing clinical narratives with respect to sentiment:

- Modelling of implicit clinical context and determining implicit sentiment
- Building upon a domain-specific sentiment lexicon
- Determining sentiment depending on the contexts
- Modelling different aspects of the patient’s status

Additional aspects include determining the opinion-holder and considering time. During treatment, the health status is supposed to improve, but can also worsen; an operation could start normally, but then become critical. For this reason, medical sentiment should also be considered over time. Time is provided by document time stamps or sometimes in the documents themselves, or clinical data could be sorted according to treatment phases. As already shown in other studies, negations are used very frequently in clinical narratives. In sentiment analysis, it is crucial to identify negations since polarity may be reversed (e.g. in *no complaints of pain when asked*). Existing algorithms such as NegFinder [38] or NegEx [39] can be exploited for this purpose.

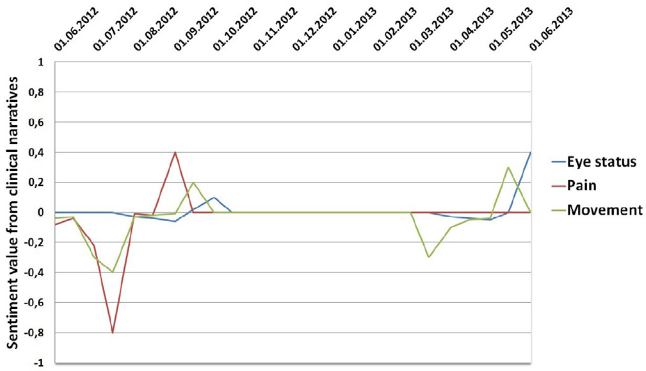

Fig. 4. Application of sentiment analysis to generate a health status graph. The lines show the status of each medical condition.

5.2. Application areas

Sentiment analysis can be applied in the medical domain in multiple ways. Most of the related work concentrated on processing medical web documents. Reported applications include:

- Mining and retrieving personal health information and opinions [31,24,22,25,23]
- Measuring quality of document content or of healthcare interactions [21,35]
- Analysing emotions and studying emotional effects [34,30,26]
- Determining clinical outcome [32]
- Detecting adverse drug events [29]

Mining and retrieving health information and opinions as well as determining clinical outcome are applications that are also relevant when performing sentiment analysis of clinical narratives. However, additional applications are possible. A few examples are presented below.

Health status aggregation. For physicians, sentiment analysis results can be used to **identify and summarize the health status** of a patient and its development over time in general or for specific medical conditions (see Fig. 4). By extracting and aggregating the attitudes and intentions expressed in clinical narratives from all the cooperating physicians an overall status from different medical perspectives can be illustrated based on the extracted opinions. Physicians diagnose the status of the patient progressively through examinations, observations and discussions with others. Each step reflects one aspect of the patient's status. Recovering the status information and personal perceptions of individual health carers is necessary to reconstruct the entire status of the patient, since it provides the connection between treatment, and the transition of treatments as well as the outcomes.

To demonstrate this usage, consider a patient with a long-term medical condition such as diabetes. From time to time, additional medical conditions related to diabetes occur. Multiple physicians are involved in treatment. However, the general practitioner receives clinical summaries from the other colleagues and needs to aggregate this information. Sentiment analysis methods could be applied to generate a curve over time showing the patient's health status as a whole. Information could also be plotted for individual medical conditions that have occurred for some time and been treated. Updated by changes to their medication (including dosage) as well as any surgery or other treatment carried out, a graph like this could provide an initial overview of the patient's current status and the development of their health. In this way, decision-making can take into account the patient's overall status reconstructed from the **"wisdom of crowds" scattered in multiple clinical narratives**.

Similarly, a graph like this could also help patients monitor their own health. When such a graph is generated from patient questionnaires or online diaries, information can be compared to clinical information to see whether the clinical impression matches the patient's perceptions.

Quality assessment. Sentiment analysis could be used to **collect statistical information on the outcome of treatments** in response to a medical condition to improve clinical decision-making and quality management. For example, it would enable analysis of how often a certain complication or critical event occurs following a specific treatment. Thus, for a hospital manager and resource planner, sentiment analysis results can provide a basic quantitative parameter based on different medical narratives which can be used as a reference for quality assessment, planning and scheduling.

Furthermore, consider patients in palliative (end-of-life) care: They require not only anaesthetic treatment to eliminate the symptoms and pain, but also the mental support of practitioners and comfort from social connections. Patient's feedback in the palliative ward is crucial to assess the effectiveness of the treatment. Based on the feedback from patients reflected in nurse notes, different aspects of treatment can be evaluated and therapy adjusted accordingly.

Outcome research. For researchers, sentiment analysis methods can be employed to extract and classify the outcome of different types of medical treatment and learn about their effectiveness. This would facilitate the labour-intensive user studies of treatment. Furthermore, sentiment analysis could be harnessed to perform a **retrospective analysis of treatment processes and outcome research**. Information about treatment outcome is captured in clinical documents and biomedical literature. An automatic analysis of patient records with respect to sentiment could help in learning more about outcomes and influence factors of diseases and treatments. Considering the individual patient's treatment, sentiment analysis methods could support in **the collection of information about the patient's health status, or the detection of critical events** that occurred during treatment. For example, information on problems such as severe haemorrhaging could be extracted from operation records. Such information might be relevant in surgical planning. Moreover, decisions on follow-up treatment should also take into consideration any problems occurring during an intervention.

Sentiment analysis could be employed to gain insights from both medical social media and clinical documents regarding the **effectiveness of a treatment or medication**. Social media provide an additional source of information previously unconsidered. They offer the opportunity to learn from patients' experiences since they describe their personal perceptions of types of treatment in social media such as drug reviews. This information augments the more implicitly described information on effectiveness in clinical narratives and biomedical literature.

6. Conclusion

This paper introduced the multiple facets of sentiment in the context of medicine and outlined areas of future research in medical sentiment analysis. While existing work in sentiment analysis from medical texts considered medical social media and biomedical literature, future research should also target the processing of clinical documents. This analysis could be used for many purposes. Uncertainty, attitudes and implicit sentiment could be collected and taken into account in clinical decision-making. The first steps towards sentiment analysis solutions could be the construction of a domain-specific sentiment lexicon and the development of methods that allow sentiment to be judged depending on the context.

Acknowledgements

This research was conducted within the research group "Digital Patient Modelling" funded by the German Ministry of Research and Education.

Appendix A. Subjectivity lexicon: Top 10 frequent terms in descending order

- **Radiology report:** right, impression, left, normal, well, advanced, effusion, free,fracture, stable
- **Discharge summary:** right, blood, stable, normal, care, illness, pain, invasive, complaint, pertinent
- **Nurse letter:** care, will, clear, vent, stable, pain, right, well, support, intact
- **Drug review:** like, will, want, right, even, good, game, need, well, better
- **MedBlog:** risk, help, need, less, disease, like, heart, better, cancer, care
- **Interview:** like, will, want, right, good, need, well, better, down, sure

Appendix B. SentiWordNet: Top 10 frequent terms in descending order

- **Radiology report:** left, artery, contrast, man, patient, small, normal, subclavian, internal, pneumothorax
- **Discharge summary:** patient, day, pain, left, normal, artery, stable, disease, discharge, edema
- **Nurse letter:** clear, pain, yellow, thick, care, shift, small, follow, good, lasix
- **Drug review:** pain, day, medicine, effects, life, feel, anxiety, medication, time, night
- **MedBlog:** risk, cancer, disease, heart, health, time, pain, high, treatment, brain
- **Interview:** game, make, good, time, point, question, work, problem, violent, system

References

[1] Hu M, Liu B. Mining and summarizing customer reviews. In: Proceedings of the 10th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '04. New York, NY, USA: ACM; 2004. p. 168–77.
[2] Bing Liu. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies. Morgan & Claypool Publishers; 2012.
[3] Balahur A, Steinberger R, Kabadjov MA, Zavarella V, Van der Goot E, Halkia M, et al. Sentiment analysis in the news. Comput Res Repos (CoRR) 2013, abs/1309.6202.
[4] Friedman C, Rindflesch TC, Corn M. Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the national library of medicine. J Biomed Inform 2013;46(5):765–73.
[5] Mishne G. Experiments with mood classification in blog posts. In: 1st Workshop on stylistic analysis of text for information access. 2005.
[6] Baldoni M, Baroglio C, Patti V, Rena P. From tags to emotions: ontology-driven sentiment analysis in the social semantic web. Intell Artif 2012;6(1):41–54.
[7] Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on empirical methods in natural language processing – Volume 10, EMNLP '02. Stroudsburg, PA, USA: Association for Computational Linguistics; 2002. p. 79–86.
[8] Pang B, Lee L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd annual meeting on association for computational linguistics, ACL '04. Stroudsburg, PA, USA: Association for Computational Linguistics; 2004.
[9] Turney PD. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics, ACL '02. Stroudsburg, PA, USA: Association for Computational Linguistics; 2002. p. 417–24.
[10] Pontiki M, Galanis D, Pavlopoulos J, Papageorgiou H, Androutsopoulos I, Manandhar S. Semeval-2014 task 4: aspect based sentiment analysis. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). August 2014. p. 27–35.
[11] Denecke K. Are sentiwordnet scores suited for multi-domain sentiment classification? Proceedings of the fourth international conference on digital information management, 2009. ICDIM 2009. 2009. p. 1–6.
[12] Montejo-Ráez A, Martínez-Cámara E, Martín-Valdivia MT, López LAUreña. Random walk weighting over sentiwordnet for sentiment polarity detection on twitter. In: Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis, WASSA '12. Stroudsburg, PA, USA: Association for Computational Linguistics; 2012. p. 3–10.
[13] Baccianella S, Esuli A, Sebastiani F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Calzolari N (Conference Chair), Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D, editors. Proceedings of the seventh international conference on language resources and evaluation (LREC'10). Valletta, Malta: European Language Resources Association (ELRA); 2010.
[14] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the conference on human language technology and empirical methods in natural language processing, HLT '05. Stroudsburg, PA, USA: Association for Computational Linguistics; 2005. p. 347–54.
[15] Valitutti R. Wordnet-affect: an affective extension of wordnet. In: Proceedings of the 4th international conference on language resources and evaluation. 2004. p. 1083–6.
[16] Mohammad SM, Turney PD. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, CAAGET '10. Stroudsburg, PA, USA: Association for Computational Linguistics; 2010. p. 26–34.
[17] Deng L, Choi Y, Wiebe J. Benefactive/malefactive event and writer attitude annotation. In: Proceedings of the 51st annual meeting of the association for computational linguistics (Volume 2: Short Papers). Sofia, Bulgaria: Association for Computational Linguistics; August 2013. p. 120–5.
[18] Goeuriot L, Na J-C, Min Kyaing WY, Khoo C, Chang Y-K, Theng Y-L, et al. Sentiment lexicons for health-related opinion mining. In: Proceedings of the 2Nd ACM SIGHIT international health informatics symposium, IHI '12. New York, NY, USA: ACM; 2012. p. 219–26.
[19] Ohana B, Tierney B, Delany S. Domain independent sentiment classification with many lexicons. In: 2011 IEEE workshops of international conference on advanced information networking and applications (WAINA). March 2011. p. 632–7.
[20] Stone PJ, Hunt EB. A computer approach to content analysis: studies using the general inquirer system. In: Proceedings of the May 21–23, 1963, spring joint computer conference, AFIPS '63 (Spring). New York, NY, USA: ACM; 1963. p. 241–56.
[21] Denecke K, Nejdl W. How valuable is medical social media data? Content analysis of the medical web. J Inf Sci 2009;179:1870–80.
[22] Ali T, Schramm D, Sokolova M, Inkpen D. Can i hear you? Sentiment analysis on medical forums. In: Proceedings of the sixth international joint conference on natural language processing. Nagoya, Japan: Asian Federation of Natural Language Processing; October 2013. p. 667–73.
[23] Xia L, Gentile AL, Munro J, Iria J. Improving patient opinion mining through multi-step classification. In: Proceedings of the 12th international conference on text, speech and dialogue, TSD '09. Berlin, Heidelberg: Springer-Verlag; 2009. p. 70–6.
[24] Na J-C, Min Kyaing WY, Khoo CSG, Foo S, Chang Y-K, Theng Y-L. Sentiment classification of drug reviews using a rule-based linguistic approach. In: Chen H-H, Chowdhury G, editors. The outreach of digital libraries: a globalized resource network, vol. 7634 of lecture notes in computer science. Springer Berlin Heidelberg; 2012. p. 189–98.
[25] Sokolova M, Matwin S, Jafer Y, Schramm D. How joe and jane tweet about their health: Mining for personal health information on twitter. In: Angelova G, Bontcheva K, Mitkov R, editors. RANLP. RANLP 2011 Organising Committee /ACL; 2013. p. 626–32.
[26] Biyani P, Caragea C, Mitra P, Zhou C, Yen J, Greer GE, et al. Co-training over domain-independent and domain-dependent features for sentiment analysis of an online cancer support community. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM '13. New York, NY, USA: ACM; 2013. p. 413–7.
[27] Ofek N, Caragea C, Rokach L, Biyani P, Mitra P, Yen J, et al. Improving sentiment analysis in an online cancer survivor community using dynamic sentiment lexicon. In: 2013 International conference on social intelligence and technology (SOCIETY). May 2013. p. 109–13.
[28] Smith P, Lee M. Cross-discourse development of supervised sentiment analysis in the clinical domain. In: Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis, WASSA '12. Stroudsburg, PA, USA: Association for Computational Linguistics; 2012. p. 79–83.
[29] Sharif H, Abbasi A, Zafar F, Zimbra D. Detecting adverse drug reactions using a sentiment classification framework. In: Proceedings of the sixth ASE international conference on social computing (SocialCom). 2014.
[30] Sokolova M, Bobicev V. What sentiments can be found in medical forums. In: Proceedings of recent advances in natural language processing. 2013. p. 633–9.
[31] Melzi S, Abdaoui A, Aze J, Bringay S, Poncelet P, Galtier F. Patients rationale: patient knowledge retrieval from health forums. In: Proceedings of the eTelemed 2014: sixth conference on eHealth, telemedicine and social medicine. 2014. p. 140–5.

[32] Niu Y, Zhu X, Li J, Hirst G. Analysis of polarity information in medical text. In: Proceedings of the American medical informatics association 2005 annual symposium. 2005. p. 570–4.

[33] Sarker A, Molla D, Paris C. Outcome polarity identification of medical papers. In: Proceedings of the Australasian language technology association workshop 2011. 2011. p. 105–14.

[34] Pestian JP, Matykiewicz P, Linn-Gust M, South B, Uzuner O, Wiebe J, et al. Sentiment analysis of suicide notes: a shared task. Biomed Inform Insights 01 2012;5:3–16.

[35] Cambria E, Benson T, Eckl C, Hussain A. Sentic proms: Application of sentic computing to the development of a novel unified framework for measuring health-care quality. Expert Syst Appl September 2012;39(12):10533–43.

[36] Liu H, Singh P. Conceptnet — a practical commonsense reasoning tool-kit. BT Technol J 2004;22(4):211–26.

[37] Cambria E, Livingstone A, Hussain A. The hourglass of emotions. In: Proceedings of the 2011 international conference on cognitive behavioural systems, COST'11. Berlin, Heidelberg: Springer-Verlag; 2012. p. 144–57.

[38] Mutalik PG, Deshpande A, Nadkarni PM. Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. J Am Med Inform Assoc JAMIA 2001;8(6): 598–609.

[39] Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001;34(5):301–10.