

Reflections on Sentiment/Opinion Analysis

Jiwei Li¹ and Eduard Hovy²

¹Computer Science Department, Stanford University, Stanford, CA 94305, USA

²Language Technology Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA

jiweil@stanford.edu hovy@cmu.edu

1 Introduction

Sentiment analysis is an application of natural language processing that focuses on identifying expressions that reflect authors' opinion-based attitude (i.e., good or bad, like or dislike) toward entities (e.g., products, topics, issues) or facets of them (e.g., price, quality).

Since the early 2000s, a large number of models and frameworks have been introduced to address this application, with emphasis on various aspects like opinion related entity extraction, review mining, topic mining, sentiment summarization, recommendation, and these extracted from significantly diverse text sources including product reviews, news articles, social media (blogs, Twitter, forum discussions), and so on.

However, despite this activity, disappointingly little has been published about what exactly a *sentiment* or *opinion* actually is. It is generally simply assumed that two (or perhaps three) polar values (*positive*, *negative*, *neutral*) are enough, and that they are clear, and that anyone would agree on how to assign such labels to arbitrary texts. Further, existing methods, despite employing increasingly sophisticated (and of course more powerful) models (e.g., neural nets), still essentially boil down to considering individual or local combinations of words and matching them against predefined lists of words with fixed sentiment values, and thus hardly transcend what was described in the early work by Pang et al. (2002).

There is nothing against simple methods when they work, but they do not always work, and without some discussion of why not, and where to go next, this application remains rather technically uninteresting. The goal of this paper is to identify gaps in the current sentiment analysis literature and to outline practical computational ways to address these issues.

Goals, Expectations and Sentiments. We begin with the fundamental question “What make people hold positive attitudes towards some entities and negative attitudes toward others?”. The answer to this question is a psychological state that relates to the opinion holder's satisfaction and dissatisfaction with some aspect of the topic in question. One of only two principal factors determines the answer: either (1) the holder's deep emotionally-driven, non-logical **native preferences**, or (2) whether (and how well) one of the holder's **goals** is fulfilled, and how (in what ways) the goal is fulfilled.

Examples of the former are reflected in sentences like “I just like red” or “seeing that makes me happy”. They are typified by adverbs like “just” and “simply” that suggest that no further conscious psychological reflection or motivation obtains. Of this class of factor we can say nothing computationally, and do not address it in the rest of this chapter.

Fortunately, a large proportion of the attitudes people write about reflect the other factor, which one can summarize as goal-driven utility. This relates primarily to Consequentialism: both to Utilitarianism, in which pleasure, economic well-being and the lack of suffering are considered desirable, but also to the general case that morally justifiable actions (and the objects that enable them) are desirable. That is, the ultimate basis for any judgment about the rightness or wrongness of one's actions, and hence of the objects that support/enable them, is a consideration of their outcome, or consequence.

In everyday life, people establish and maintain goals or expectations, both long-term or short-term, urgent or not-urgent, ones. Achieving these goals would fill one with satisfaction, otherwise dissatisfaction: a man walks into a restaurant to achieve the goal of getting full, he cannot be satisfied if all food was sold out (the main goal not

being achieved). A voter would not be satisfied if his candidate or party fails to win an election, since the longer-term consequences would generally work against his own preferences. The generation of sentiment-related texts is guided by such sorts of mental satisfaction and dissatisfaction induced by goals being achieved or needs being fulfilled.

We next provide some examples to illustrate why identifying these aspects is essential and fundamental for adequate sentiment/opinion analysis. Following the most popular motivation for computational sentiment analysis, suppose we wish to analyze customers' opinions towards a product or an offering. It is not sufficient to simply determine that someone likes or dislikes something; to make that knowledge useful and actionable, one also wants to know *why* that is the case. Especially when one would like to change the opinion, it is important to determine what it is about the topic that needs to be changed.

Case (1)

- **Question:** *Why did the customer like detergent X?*
- **Customer's review:** *The detergent removes stubborn stains.*

No general sentiment indicator is found in the above review. But the review directly provides the reason, and assuming his/her goal of clean clothing is achieved, it is evident that the opinion holder holds a positive opinion towards the detergent.

Case (2)

- **Question:** *Why did the traveller dislike flight Y?*
- **Customer's review:** *The food was good. The crew was helpful and took care of everything. The service was efficient. However the flight was supposed to take 1.5 hours but was 3 hours late, and I missed my next connecting flight.*

The major goal of taking a flight is to get to your destination, which is more important than goals like enjoying one's food and receiving pampering service. While multiple simultaneous goals induce competing opinion decisions, the presence of an importance ranking among them determines the overall sentiment.

Case (3)

- **Question:** *Why did the customer visit restaurant Z?*
- **Review1:** *The food is bad.*
- **Review2:** *The waiter was kind but the food was bad.*
- **Review3:** *The food was good but the waiter was rude.*

Although the primary goal of being sated may be achieved, secondary goals such as enjoying the food and receiving respectful service can be violated in various combinations. Often, these goals pertain to the method by which the primary goal was achieved; in other words, to the question "how?" rather than "why?".

A sentiment determination algorithm that can provide more than just a simple opinion label thus has to pay attention both to the primary reason behind the holder's involvement with the topic ("why?") and to the secondary reasons (both "why?" and "how?"), and has to be able to determine their relative importance and relationship to the primary goal.

Goals and Expectations are Personal. As different people (opinion holders) are from different backgrounds, have different personalities, and are in different situations, they have different goals, needs, and the expectations of life. This diversity generally leads to completely diverse opinions towards the same entity, the same action, and the same situation: a billionaire wouldn't be the least bit concerned with the price in a bread shop but would consider the quality, while a beggar might care only about the price. This rather banal observation is explained best by Maslow's famous hierarchy of needs (Maslow, 1943), in which the beggar's attention focuses on Maslow's Physiological needs while the billionaire's focuses on Self-Actualization; more on this in Section 3.1.

Life Requires Trade-offs. Most situations in real life address many personal needs simultaneously. People thus face trade-offs between their goals, which entails sacrificing the achievement of one goal for the satisfaction of another. Given the variability among people, the rankings and decision procedures will also from individual to individual. However, Maslow's hierarchy describes

the general behavioral trends of people in most societies and situations.

Complex Sentiment Expressions. As far as we see, current opinion analysis frameworks mostly fail to address the kinds of issues mentioned above, and thereby impair a deeper understanding about opinion or sentiment. As a result, they find it impossible to provide even rudimentary approaches to cases such as the following (from (Hovy, 2015)):

1. *Peter thinks the pants are great and I cannot agree more.*
2. *Peter thinks the pants are great but I don't agree.*
3. *Sometime I like it but sometimes I hate it.*
4. *He was half excited, half terrified.*
5. *The movie is indeed wonderful, but for some reason, I just don't like it.*
6. *Why I won't buy this game even though I like it.*

In this paper, we explore the feasibility of addressing these issues in a practical way using machine learning techniques currently available.

2 A Review of Current Sentiment Analysis

Here we give a brief overview of tasks in current sentiment analysis literature. More details can be found in (Liu, 2010; Liu, 2012).

The key points involved at the algorithm level in the sentiment analysis literature follow the basic approaches of statistical machine learning, in which a gold-standard labeling of training data is obtained through manual annotation or other data harvesting approaches (e.g., semi-supervised or weakly supervised), and this is then used to train a variety of association-learning techniques who are then tested on new material. Usually, some text unit has to be identified and then associated with a sentiment label (e.g., positive, neutral, negative). Based on the annotated dataset, the techniques learn that vocabulary items like “bad”, “awful”, and “disgusting” are negative sentiment indicators while “good”, “fantastic” and “awesome” are positive ones. The main complexity lies in learning which words carry some opinion and, especially,

what to decide in cases where different words with opposite labels appear in the same clause.

Basic sentiment analysis identifies the simple polarity of a text unit (e.g., a token, a phrase, a sentence, or a document) and is framed as a binary or multi-class classification task; see for example Pang et al.'s work (2002) that uses a unigram/bigram feature-based SVM classifier. Over the past 15 years, techniques have evolved from simple rule-based word matching to more sophisticated feature and signal (e.g., local word composition, facets of topics, opinion holder) identification and combination, from the level of single tokens to entire documents, and from ‘flat’ word strings without any syntactic structure at all to incorporation of complex linguistic structures (e.g., discourse or mixed-affect sentences); see (Pang and Lee, 2004; Hu and Liu, 2004; Wiebe et al., 2005; Nakagawa et al., 2010; Maas et al., 2011; Tang et al., b; Qiu et al., 2011; Wang and Manning, 2012; Tang et al., a; Yang and Cardie, 2014a; Snyder and Barzilay, 2007). Recent progress in neural models provides new techniques for local composition of both opinion and structure (e.g., subordination, conjunction) using distributed representations of text units (e.g., (Socher et al., 2013; Irsoy and Cardie, 2014a; Irsoy and Cardie, 2014b; Tang, 2015; Tang et al., 2014)).

A supporting line of research extends the basic sentiment classification to include related aspects and facets, such as identifying opinion holders, the topics of opinions, topics not explicitly mentioned in the text, etc.; see (Choi et al., 2006; Kim and Hovy, 2006; Kim and Hovy, 2004; Li and Hovy, 2014; Jin et al., 2009; Breck et al., 2007; Johansson and Moschitti, 2010; Yang and Cardie, 2012; Yang and Cardie, 2013; Yang and Cardie, 2014b). These approaches usually employ sequence labeling models (e.g., CRF (Lafferty et al., 2001), HMM (LIU et al., 2004)) to identify whether the current token corresponds to a specific sentiment-related aspect or facet.

An important part of such supportive work is the identification of the relevant aspects or facets of the topic (e.g., the ambience of a restaurant vs. its food or staff or cleanliness) and the correspondent sentiment; see (Brody and Elhadad, 2010; Lu et al., 2011; Titov and McDonald, 2008; Jo and Oh, 2011; Xueke et al., 2013; Kim et al., 2013; García-Moya et al., 2013; Wang et al., 2011; Moghaddam and Ester, 2012). Online reviews (about products

or offerings) in crowdsourcing and traditional sites (e.g., yelp, Amazon, Consumer Reports) include some sort of aspect-oriented star rating systems where more stars indicate higher level of satisfaction. Consumers rely on these user-generated on-line reviews when making purchase decisions. To tackle this issue, researchers invent aspect identification or target extraction approaches as one sub-field of sentiment analysis. These approaches first identify 'aspects/facets' of the principal Topic and then discover authors' corresponding opinions for each one; e.g., (Brody and Elhadad, 2010; Titov and McDonald, 2008). Aspects are usually identified either manually or automatically using word clustering models (e.g., LDA (Blei et al., 2003) or pLSA). However, real life is usually a lot more complex and much harder to break into a series of facets (e.g., quality of living, marriage, career).

Other related work includes opinion summarization, aiming to summary sentiment key points given long texts (e.g., (Hu and Liu, 2004; Liu et al., 2005; Zhuang et al., 2006; Ku et al., 2006)), opinion spam detection aiming at identifying fictitious reviews generated to deceive readers (e.g., (Ott et al., 2011; Li et al., ; Li et al., 2013; Jindal and Liu, 2008; Lim et al., 2010)), sentiment text generation (e.g., (Mohammad, 2011; Blair-Goldensohn et al., 2008)), and large-scale sentiment/mood analysis on social media for trend detection (e.g., (O'Connor et al., 2010; Bollen et al., 2011; Conover et al., 2011; Paul and Dredze, 2011)).

3 The Needs and Goals behind Sentiments

As outlined in Section 1, this chapter argues that an adequate and complete account of utilitarian-based sentiment is possible only with reference to the goals of the opinion holder. In this section we discuss a classic model of human needs and associated goals and then outline a method for determining such goals from text.

3.1 Maslow's Hierarchy of Needs

Abraham Maslow (Maslow, 1943; Maslow, 1967; Maslow et al., 1970; Maslow, 1972) developed a theory of the basic human needs as being organized in a hierarchy of importance, visualized using a pyramid (shown in Figure 1), where needs at the bottom are the most pressing, basic, and fundamental to human life (that is, the human will tend

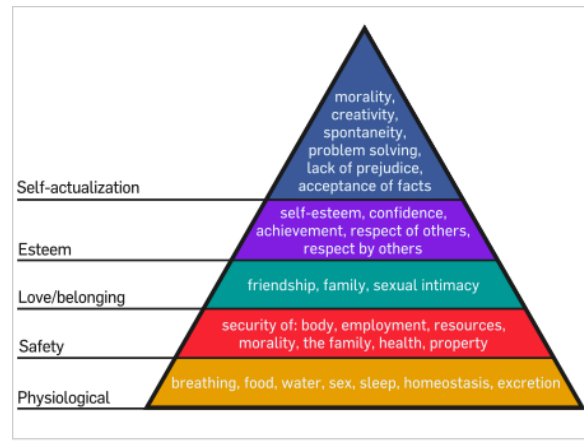


Figure 1: Maslow's Hierarchy of Needs. Figure borrowed from Wikipedia https://en.wikipedia.org/wiki/Abraham_Maslow

to choose to satisfy them first before progressing to needs higher up).

According to Maslow's theory, the most basic two levels of human needs are¹:

- Physiological needs: breathing, food, water, sleep, sex, excretion, etc.
- Safety Needs: security of body, employment, property, health, etc.

which are essential for the physical survival of a person. Once these needs are satisfied, people tend to accomplish more and move to higher levels:

- Love and Belonging: psychological needs like friendship, family, sexual intimacy.
- Esteem: the need to be competent and recognized such as through status and level of success like achievement, respect by others, etc.

These four types of needs are also referred to as DEFICIT NEEDS (or D-NEEDS), meaning that for any human, if he or she doesn't have enough of any of them, he or she will experience the desire to obtain them. Less pressing than the D-needs are the so-called GROWTH NEEDS, including Cognitive, Aesthetic (need for harmony, order and beauty), and Self-actualization (described

¹References from https://en.wikipedia.org/wiki/Abraham_Maslow; https://en.wikipedia.org/wiki/Maslow's_hierarchy_of_needs; <http://www.edpsycinteractive.org/topics/conation/maslow.html>

by Maslow as “the desire to accomplish everything that one can, to become the most that one can be”). Growth needs are more generalized, obscure, and computationally challenging. We focus in this chapter on deficit needs. For further reading, refer to Maslow’s original papers (1943; 1967) or relevant Wikipedia pages.

We note that real life offers many situations in which an action does not easily align with a need listed in the hierarchy (for example, the goal of British troops to arrest an Irish Republican Army leader or of US troops to attack Iraq). Additionally, a single action (e.g., going to college, looking for a job) can simultaneously address multiple needs. Putting aside such complex situations in this chapter, we focus on more tractable situations to illustrate the key points².

3.2 Finding Appropriate Goals for Actions and Entities

Typically, each deficit need gives rise to one or more goals that impel the agent (the opinion holder) to appropriate action. Following standard AI and Cognitive Science practice, we assume that the agent instantiates one or more plans to achieve his or her goals, where a plan is a sequence of actions intended to alter the state of the world from some situation (typically, the agent’s initial state) to a situation in which the goal has been achieved and the need satisfied. In each plan, its actions, their preconditions, and the entities used in performing them (the plan’s so-called *props*) constitute the material upon which sentiment analysis operates. For example, the goal to *sate one’s hunger* may be achieved by plans such as *visit-restaurant*, *cook-and-eat-meal-at-home*, *buy-or-steal-ready-made-food*, *cadge-meal-invitation*, etc. In all these plans, *food* is one of the props. For the restaurant and buying-food plans, an affordable *price* is an important precondition.

A sentiment detection system that seeks to understand why the holder holds a specific opinion valence has to determine the specific actions, preconditions, and props that are relevant to the holder’s goal, and to what degree they suffice. In principle, a complete account requires the system to infer from the given text:

²However, putting them aside doesn’t mean that we don’t need to explore and explain these complex situations. On the contrary, these situations are essential and fundamental to the understanding of opinion and sentiment, but requires deeper and more systematic exploration in psychology, cognitive science, and AI.

1. what need is active,
2. which goal(s) have been activated to address the need,
3. which plan(s) is/are being followed to achieve the goal(s),
4. which actions, preconditions, and props appear in these plan(s),
5. which of these is/are being talked about in the text,
6. how well it/they actually have furthered the agent’s plan(s),

from which the sentiment valence can be automatically deduced. When the valence is given in the text, one can work ‘backwards’ to infer step 6, and possibly even earlier steps.

Determining all this is a tall order for computational systems. Fortunately, it is possible to circumvent much of this reasoning in practice. For most common situations, a relatively small set of goals and plans obtains, and the relevant actions, preconditions, and props are usually quite standard. (In fact, they are precisely what is typically called ‘facets’ in the sentiment analysis literature, for which, as described in Section 2, various techniques have been investigated, albeit without a clear understanding of the reason these facets are important.)

Given this, the principal unaddressed computational problem today is the determination from the text of the original need or goal being experienced by the holder, since that is what ties together all the other (and currently investigated) aspects. How can one, for a given topic, determine the goals an agent would typically have for it, suggest likely plans, and potentially pinpoint specific actions, preconditions, and props?

One approach is to perform automated goal and plan harvesting, using typical text mining / pattern-matching approaches from Information Extraction. This is a relatively mature application of NLP (Hearst, 1992; Riloff, 1997; Riloff, 1999; Snow, 2005; Davidov, 2006; Etzioni, 2005; Banko, 2009; Mitchell, 2009; Ritter, 2009 title-What is this anyway Automatic hypernym discovery booktitleProceedings of the AAAI spring symposium on learning by reading and learning to

read; Kozareva and Hovy, 2013), and the harvesting power and behavior of various styles of patterns has been investigated for over two decades. (In practice, the Double-Anchored Pattern (DAP) method (Kozareva and Hovy, 2013) works better than most others.) Stated simply, one creates or automatically induces text patterns anchored on the topic (e.g., a camera) such as

“I want a camera because * ”
“If I had a camera I could * ”
“the main reason to get a camera is * ”
“wanted to *, so he bought a camera”
etc.

and then extracts from large amounts of text the matched VPs and NPs as being relevant to the topic. Appropriately rephrased and categorized, one obtains the information harvested by these patterns would provide typical goals (reasons) for buying and using cameras.

4 Toward a Practical Computational Approach

We are now ready to describe the overall approach necessary for a more complete sentiment analysis system. For illustrative purposes we focus on simple binary (positive/negative) valence identification. However, the framework applies to finer granularity (e.g., multi-class classification, regression) with minor adjustments. We first provide an overall algorithm sketch, provide a series of examples, and then suggest models for determining the still unexplored aspects required for deeper sentiment analysis.

First, we assume that standard techniques are employed to find the following from some given text:

1. Opinion Holder: Individual or organization holding the opinion.
2. Entity/Aspect/Theme/Facet: topic or aspect about which the opinion is held.
3. Sentiment Indicator: Sentiment-related text (tokens, phrases, sentences, etc.) that indicate the polarity of the holder.
4. Valence: *like*, *neutral*, or *dislike*.

These have been defined (or at least used with implicit definition) throughout the sentiment literature, and are defined for example in (Hovy, 2015).

Of these, item 1 is usually achieved by simple matching. Item 2 can be partially addressed by recent topic/facet mining models, and item 3 can be addressed by existing sentiment related algorithms at the word-, sentence-, or text-level. Item 4 at its simplest is a matter of keyword matching, but the composition within a sentence of contrasting valences has generated some interesting research. Annotated corpora (or other semi-supervised data harvesting techniques) might be needed for goal and need identification, as discussed above.

Given this, the following sketch algorithm implements deeper sentiment analysis:

-
1. In the text, identify the key goal underlying the Theme.
 2. Is there is no apparent goal?
 - If yes, the opinion is probably non-utilitarian, so find and return a valence if any, but return no reason for it.
 - If no, go to step 3.
 3. Determine whether the goal is satisfied:
 - If yes, go to step 4,
 - If no, return a negative valence.
 4. Identify the subgoals involved in achieving the major goal.
 5. Identify how well the subgoals are satisfied.
 6. Determine the final utilitarian sentiment based on the trade-off between different subgoals, and return it together with the trade-off analysis as the reasoning.

This procedure requires the determination of the Goals or Subgoals and the Condition/Situation under which the opinion holder holds that opinion. The former is discussed above; the latter can usually be determined from the context of the given text.

4.1 Examples and Illustration

As a running example we use simple restaurant reviews, sentences in italics indicating original text from the reviews³:

Case 1

1. *My friends and I went to restaurant X.*
2. *So many people were waiting there and we left without eating.*

³These reviews were originally from yelp reviews and revised by the authors for illustration purposes.

Following the algorithm sketch, the question “**was the major goal of going to a restaurant fulfilled?**” is answered **no**. The reviewer is predicted to hold a negative sentiment. Similar reasoning applies to Case 2 in Section 1.

Case 2

1. *My friends and I went to restaurant X.*
2. *The waiter was friendly and knowledgeable.*
3. *We ordered curry chicken, potato chips and italian sausage. The Italian sausage was delicious.*
4. *Overall the food was appetizing,*
5. *but I just didn't enjoy the experience.*

To the question “**was the major goal of being full fulfilled?**” the answer is **yes**, as the food was ordered and eaten. Next the algorithm addresses the *how* (manner of achievement) question described in steps 4–6, which involves the functional elements of goals/needs embedded in each sentence:

1. *My friends and I went to restaurant X.*
Opinion Holder: I
Entity/Aspect/Theme: restaurant X
Need: sate hunger
Goal: visit restaurant
Sentiment Indicator: none
Valence: neutral Condition: in restaurant X
2. *The waiter was friendly and knowledgeable.*
Opinion Holder: I
Entity/Aspect/Theme: waiter
Need: gather respect/friendship
Subgoal: order food
Sentiment Indicator: friendly, knowledgeable
Valence: positive
Condition: in restaurant X
3. *We ordered curry chicken, potato chips and italian sausage. Italian sausage was delicious.*
Opinion Holder: I
Entity/Aspect/Theme: Italian sausage
Need: sate hunger
Subgoal: eat food
Sentiment Indicator: delicious
Valence: positive
Condition: in restaurant X

4. *Overall the food was appetizing,*
Opinion Holder: I
Entity/Aspect/Theme: food
Need: sate hunger
Subgoal: eat enough to remove hunger
Sentiment Indicator: appetizing
Valence: positive
Condition: in restaurant X
5. *but I just didn't enjoy the experience.*
Opinion Holder: I
Entity/Aspect/Theme: restaurant visit experience
Need: none — this is not utilitarian
Goal: none
Sentiment Indicator: didn't enjoy
Sentiment Label: negative
Condition: in restaurant X

The analysis of the needs/goals and their respective positive and negative valences allows one to justify the various sentiment statements, and (in the case of the final negative decision) also indicate that it is not based on utilitarian considerations.

4.2 A Computational Model of Each Part

Current computational models can be used to address each of the aspects involved in the sketch algorithm. We provide only a high-level description of each.

Deciding Functional Elements. Case 2 above involves three of the needs described in Maslow's hierarchy: food, respect/friendship, and emotion. The first two are stated to have been achieved. The third is a pure emotion, expressed without a reason, why the holder “just didn't enjoy the experience”. Pure emotions usually have no overt utilitarian value but only relate to the holder's high-level goal of being happy. In this example, we have to conclude that since all overt goals were met, either some unstated utilitarian Maslow-type need was not met, or the holder's opinion stems from a deeper psychological/emotional bias, of the kind mentioned in Section 1, that goes beyond utilitarian value.

Whether the Major Goal is Achieved. To make a decision about goal achievement, one must: (1) identify the goal/subgoal of an action (e.g., buying the detergent, going to a restaurant); (2) identify whether that goal/subgoal is achieved. The two steps can be computed either separately

or jointly using current machine learning models and techniques, including:

- **Joint Model:** Annotate corpora for satisfaction or not for all goals and subgoals together, and train a single machine learning algorithm.
- **Separate Model:**
 1. Determine the goal and its plans and subgoals either through annotation or as described in Section 3.2.
 2. Associate the actions or entities of the Theme (e.g., going to a restaurant; buying a car) with their respective (sub)goals.
 3. Align each subgoal with indicator sentence(s) in the document (e.g., “I got a small portion”; “the car was all it was supposed to be”).
 4. Decide whether the subgoal is satisfied based on indicator sentence(s).

Learning Weights for Different Goals/Needs.

One can clearly infer that the customer in case 2 assigns more weight to the emotional aspect, that being his or her final conclusion, and less to the food or respect/friendship (which comes last in this scenario). More formally, for a given text D , we discover L needs/(sub)goals, with indices $1, 2, \dots, L$. Each type of need/(sub)goal $i \in [1, L]$ is associated with a weight that contributes to the final sentiment valence decision v_i . In document D , each type of need i is associated with achievement value a_i that indicates how the need or goal is satisfied. The sentiment score S_D for given document D is then given by:

$$S_D = \sum_{i \in [1, L]} v_i \cdot a_i$$

This simple approach is comparable to a regression model that assigns weights to relevant aspects, where gold standard examples can be the overall ratings of the labeled restaurant reviews. One can view such a weight decision procedure as a supervised regression model by assigning a weight value to each discovered need. Such a procedure is similar to latent aspect rating introduced in (Wang et al., 2011; Zhao et al., 2010) by learning aspect weight (i.e., value, room, location, or service) for hotel review ratings. A simple illustrative example might be collaborative filtering

in recommendation systems, e.g., (Breese et al., 1998; Sarwar et al., 2001), optimizing need weight regarding each respective individual (which could be sampled from a uniform prior for humans’ generally accepted weights).

Since individual expectations can differ, it would be advantageous to maintain opinion holder profiles (for example, both yelp and Amazon keep individual profiles for each customer) that record one’s long-term activity. This would support individual analysis of background, personality, or social identity, and enable learning of specific goal weights for different individuals.

When these issues have been addressed, one can start asking deeper questions like:

- *Q: Why does John like his current job though his salary is low?*
A: He weighs employment more highly than family.
- *Q: How wealthy is a particular opinion holder?*
A: He might be rich as he places little concern (weight) on money.

or make user-oriented recommendations like:

- *Q: Should the system recommend an expensive-but-luxurious hotel or a cheap-but-poor hotel?*

4.3 Prior / Default Knowledge about Opinion Holders

Sentiment/opinion analysis can be considerably assisted by the existence of a knowledge base that provides information about the typical preferences of the holder.

Individuals’ goals vary across backgrounds, ages, nationalities, genders, etc. An engineer would have different life goals from a businessman, or a doctor, a citizen living in South America would have different weighing systems from those in Europe or the United States, people in wartime would have different life expectations from when in peacetime. Two general methods exist today for practically collecting such standardized knowledge to construct a relevant knowledge base:

- (1) **Rule-based Approaches.** Hierarchies of personality profiles have been proposed, and changes to them have long been explored in

the social and developmental psychology literature, usually based on polls or surveys. For example, (1981) found that children have higher physical needs than other age groups, love needs emerging in the transitional period from childhood to adulthood; esteem needs are the highest among adolescents; the highest self-actualization levels are found with adults; and the highest levels of security are found at older ages. As another example, researchers (Tang and Ibrahim, 1998; Tang et al., 2002; Tang and West, 1997) have found that survival (i.e., physiological and safety) needs dominate during wartime while psychological needs (i.e., love, self-esteem, and self-actualization) surface during peacetime, which is in line with our expectations. For computational implementation, however, these sorts of studies provide very limited evidence, since only a few aspects are typically explored.

(2) Computational Inference Approaches.

Despite the lack of information about individuals, reasonable preferences can be inferred from other resources such as online social media. A vast section of the Social Network Analysis research focuses on this problem, as well as much of the research of the large web search engine companies. Networking websites like Facebook, LinkedIn, and Google Plus provide rich repositories of personal information about individual attributes such as education, employment, nationality, religion, likes and dislikes, etc. Additionally, online posts usually offer direct evidence for such attributes. Some examples include age (Rao et al., 2010; Rao and Yarowsky, 2010), gender (Ciot et al., 2013), living location (Sadilek et al., 2012), and education (Mislove et al., 2010).

5 Conclusion and Discussion

The past 15 years has witnessed significant performance improvements in training machine learning algorithms for the sentiment/opinion identification application. But little progress has been made toward a deeper understanding about what opinions or sentiments are, why people hold them, and why and how their facets are chosen and expressed. No-one can deny the unprecedented contributions of statistical learning algorithms in

modern-day (post-1990s) NLP, for this application as for others. However, ignoring cognitive and psychological perspectives in favor of engineering alone inevitably hampers progress once the algorithms asymptote to their optimal performance, since understanding *how* to do something doesn't necessarily lead to better insight about *what* needs to be done, or how it is best represented. For example, when inter-annotator agreement on sentiment labels peaks at 0.79 even for the rather crude 3-way sentiment granularity of positive/neutral/negative (Ogneva, 2010), is that the theoretical best that could be achieved? How could one ever know, without understanding what other aspects of sentiment/opinion are pertinent and investigating whether they could constrain the annotation task and help boost annotation agreement?

In this paper, we described possible directions for deeper understanding, helping bridge the gap between psychology / cognitive science and computational approaches. We focus on the opinion holder's underlying needs and their resultant goals, which, in a utilitarian model of sentiment, provides the basis for explaining the reason a sentiment valence is held. (The complementary non-utilitarian, purely intuitive preference-based basis for some sentiment decisions is a topic requiring altogether different treatment.) While these thoughts are still immature, scattered, unstructured, and even imaginary, we believe that these perspectives might suggest fruitful avenues for various kinds of future work.

References

- Michelle Banko. 2009. Ph.D. Dissertation, University of Washington.
- Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A Reis, and Jeff Reynar. 2008. Building a sentiment summarizer for local service reviews. In *WWW Workshop on NLP in the Information Explosion Era*, volume 14.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.
- Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.
- Eric Breck, Yejin Choi, and Claire Cardie. 2007. Identifying expressions of opinion in context. In *IJCAI*.

- John S Breese, David Heckerman, and Carl Kadie. 1998. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc.
- Samuel Brody and Noemie Elhadad. 2010. An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 804–812. Association for Computational Linguistics.
- Yejin Choi, Eric Breck, and Claire Cardie. 2006. Joint extraction of entities and relations for opinion recognition. In *EMNLP*.
- Morgane Ciot, Morgan Sonderegger, and Derek Ruths. 2013. Gender inference of twitter users in non-english contexts. In *EMNLP*, pages 1136–1145.
- Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *ICWSM*.
- A. Davidov, D. & Rappoport. 2006. Efficient unsupervised discovery of word categories using symmetric patterns and high frequency words. In *Proceedings of the 21st international conference on computational linguistics COLING and the 44th annual meeting of the ACL*, pages 297–304.
- Cafarella M. Downey D. Popescu A. M. Shaked T. Soderland S. et al. Etzioni, O. 2005. Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1):91–134.
- Lisette García-Moya, Henry Anaya-Sánchez, and Rafael Berlanga-Llavori. 2013. Retrieving product features and opinions from customer reviews. *IEEE Intelligent Systems*, 28(3):0019–27.
- Barbara L Goebel and Delores R Brown. 1981. Age differences in motivation related to maslow’s need hierarchy. *Developmental Psychology*, 17(6):809.
- Marti Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational Linguistics*, pages 539–545.
- Eduard H Hovy. 2015. What are sentiment, affect, and emotion? applying the methodology of michael zock to sentiment analysis. In *Language Production, Cognition, and the Lexicon*, pages 13–24. Springer.
- Minqing Hu and Bing Liu. 2004. Mining opinion features in customer reviews. In *AAAI*, volume 4, pages 755–760.
- Ozan Irsoy and Claire Cardie. 2014a. Deep recursive neural networks for compositionality in language. In *Advances in Neural Information Processing Systems*, pages 2096–2104.
- Ozan Irsoy and Claire Cardie. 2014b. Opinion mining with deep recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 720–728.
- Wei Jin, Hung Hay Ho, and Rohini K Srihari. 2009. A novel lexicalized hmm-based learning framework for web opinion mining. In *ICML*.
- Nitin Jindal and Bing Liu. 2008. Opinion spam and analysis. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 219–230. ACM.
- Yohan Jo and Alice H Oh. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 815–824. ACM.
- Richard Johansson and Alessandro Moschitti. 2010. Syntactic and semantic structure for opinion expression detection. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1367. Association for Computational Linguistics.
- Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*.
- Suin Kim, Jianwen Zhang, Zheng Chen, Alice H Oh, and Shixia Liu. 2013. A hierarchical aspect-sentiment model for online reviews. In *AAAI*.
- Z. Kozareva and E.H Hovy. 2013. Tailoring the automated construction of large-scale taxonomies using the web. *Journal of Language Resources and Evaluation*, 47:859–890.
- Lun-Wei Ku, Yu-Ting Liang, and Hsin-Hsi Chen. 2006. Opinion extraction, summarization and tracking in news and blog corpora. In *AAAI spring symposium: Computational approaches to analyzing weblogs*, volume 100107.
- John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
- Jiwei Li and Eduard Hovy. 2014. Sentiment analysis on the people’s daily.

- Jiwei Li, Myle Ott, Claire Cardie, and Eduard Hovy. Towards a general rule for identifying deceptive opinion spam.
- Jiwei Li, Myle Ott, and Claire Cardie. 2013. Identifying manipulated offerings on review portals. In *EMNLP*, pages 1933–1942.
- Ye-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, and Hady Wirawan Lauw. 2010. Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 939–948. ACM.
- Yun-zhong LIU, Ya-ping LIN, and Zhi-ping CHEN. 2004. Text information extraction based on hidden markov model [j]. *Acta Simulata Systematica Sinica*.
- Bing Liu, Mingqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pages 342–351. ACM.
- Bing Liu. 2010. Sentiment analysis and subjectivity. *Handbook of Natural Language Processing*, 2:627–666.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Bin Lu, Myle Ott, Claire Cardie, and Benjamin K Tsou. 2011. Multi-aspect sentiment analysis with topic models. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 81–88. IEEE.
- Andrew L Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 142–150. Association for Computational Linguistics.
- Abraham Harold Maslow, Robert Frager, James Fadiman, Cynthia McReynolds, and Ruth Cox. 1970. *Motivation and personality*, volume 2. Harper & Row New York.
- Abraham Harold Maslow. 1943. A theory of human motivation. *Psychological review*, 50(4):370.
- Abraham H Maslow. 1967. A theory of metamotivation: The biological rooting of the value-life. *Journal of Humanistic Psychology*.
- Abraham H Maslow. 1972. *The Farther Reaches of Human Nature*. Maurice Bassett.
- Alan Mislove, Bimal Viswanath, Krishna P Gummadi, and Peter Druschel. 2010. You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 251–260. ACM.
- Betteridge J. Carlson A. Hruschka E. & Wang R. Mitchell, T. M. 2009. Populating the semantic web by macro-reading internet text. In *Proceedings of the 8th international semantic web conference (ISWC)*.
- Samaneh Moghaddam and Martin Ester. 2012. On the design of lda models for aspect-based opinion mining. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 803–812. ACM.
- Saif Mohammad. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114. Association for Computational Linguistics.
- Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi. 2010. Dependency tree-based sentiment classification using crfs with hidden variables. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 786–794. Association for Computational Linguistics.
- Brendan O’Connor, Ramnath Balasubramanian, Bryan R Routledge, and Noah A Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11:122–129.
- Maria Ogneva. 2010. How companies can use sentiment analysis to improve their business. Retrieved August, 30.
- Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 309–319. Association for Computational Linguistics.
- Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271. Association for Computational Linguistics.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.
- Michael J Paul and Mark Dredze. 2011. You are what you tweet: Analyzing twitter for public health. In *ICWSM*, pages 265–272.

- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27.
- Delip Rao and David Yarowsky. 2010. Detecting latent user properties in social media. In *Proc. of the NIPS MLSN Workshop*.
- Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. 2010. Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, pages 37–44. ACM.
- J. Riloff, E. & Shepherd. 1997. A corpus-based approach for building semantic lexicons. In *Proceedings of the second conference on empirical methods in natural language processing (EMNLP)*, pages 117–124.
- R. Riloff, E. & Jones. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of the sixteenth national conference on artificial intelligence (AAAI)*, pages 474–479.
- Soderland S. & Etzioni O. Ritter, A. 2009), title=What is this, anyway: Automatic hypernym discovery, booktitle=Proceedings of the AAAI spring symposium on learning by reading and learning to read.
- Adam Sadilek, Henry Kautz, and Jeffrey P Bigham. 2012. Finding your friends and following them to where you are. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 723–732. ACM.
- Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM.
- Jurafsky D. & Ng A.Y. Snow, R. 2005. Learning syntactic patterns for automatic hypernym discovery. In & L. Bottou (eds.) L. K. Saul, Y. Weiss, editor, *Advances in neural information processing systems*, volume 17, pages 1297–1304.
- Benjamin Snyder and Regina Barzilay. 2007. Multiple aspect ranking using the good grief algorithm. In *HLT-NAACL*, pages 300–307.
- Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, volume 1631, page 1642. Citeseer.
- Thomas Li-Ping Tang and Abdul H Safwat Ibrahim. 1998. Importance of human needs during retrospective peacetime and the persian gulf war: Mideastern employees. *International Journal of Stress Management*, 5(1):25–37.
- Thomas Li-Ping Tang and W Beryl West. 1997. The importance of human needs during peacetime, retrospective peacetime, and the persian gulf war. *International Journal of Stress Management*, 4(1):47–62.
- Duyu Tang, Furu Wei, Bing Qin, Li Dong, Ting Liu, and Ming Zhou. A joint segmentation and classification framework for sentiment analysis.
- Duyu Tang, Furu Wei, Bing Qin, Ming Zhou, and Ting Liu. Building large-scale twitter-specific sentiment lexicon: A representation learning approach.
- TLP Tang, AHS Ibrahim, and WB West. 2002. Effects of war-related stress on the satisfaction of human needs: The united states and the middle east. *International Journal of Management Theory and Practices*, 3(1):35–53.
- Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 1555–1565.
- Duyu Tang. 2015. Sentiment-specific representation learning for document-level sentiment analysis. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 447–452. ACM.
- Ivan Titov and Ryan T McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *ACL*, volume 8, pages 308–316. Citeseer.
- Sida Wang and Christopher D Manning. 2012. Baselines and bigrams: Simple, good sentiment and topic classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pages 90–94. Association for Computational Linguistics.
- Hongning Wang, Yue Lu, and ChengXiang Zhai. 2011. Latent aspect rating analysis without aspect keyword supervision. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 618–626. ACM.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210.
- Xu Xueke, Cheng Xueqi, Tan Songbo, Liu Yue, and Shen Huawei. 2013. Aspect-level opinion mining of online customer reviews. *Communications, China*, 10(3):25–41.
- Bishan Yang and Claire Cardie. 2012. Extracting opinion expressions with semi-markov conditional random fields. In *EMNLP*.
- Bishan Yang and Claire Cardie. 2013. Joint inference for fine-grained opinion extraction. In *ACL (1)*, pages 1640–1649.

- Bishan Yang and Claire Cardie. 2014a. Context-aware learning for sentence-level sentiment analysis with posterior regularization. In *Proceedings of ACL*.
- Bishan Yang and Claire Cardie. 2014b. Joint modeling of opinion expression extraction and attribute classification. *Transactions of the Association for Computational Linguistics*, 2:505–516.
- Wayne Xin Zhao, Jing Jiang, Hongfei Yan, and Xiaoming Li. 2010. Jointly modeling aspects and opinions with a maxent-lda hybrid. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 56–65. Association for Computational Linguistics.
- Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 43–50. ACM.