Polarity Consistency Checking for Sentiment Dictionaries

Abstract

Polarity classification of words is important for applications such as Opinion Mining and Sentiment Analysis. A number of sentiment word/sense dictionaries have been manually or (semi)automatically constructed. The dictionaries have substantial inaccuracies. Besides obvious instances, where the same word appears with different polarities in different dictionaries, the dictionaries exhibit complex cases, which cannot be detected by mere manual inspection. We introduce the concept of polarity consistency of words/senses in sentiment dictionaries in this paper. We show that the consistency problem is NP-complete. We reduce the polarity consistency problem to the satisfiability problem and utilize a fast SAT solver to detect inconsistencies in a sentiment dictionary. We perform experiments on four sentiment dictionaries and WordNet.

1 Introduction

The opinions expressed in various Web and media outlets (e.g., blogs, newspapers) are an important yardstick for the success of a product or a government policy. For instance, a product with consistently good reviews is likely to sell well. The general approach is to summarize the *semantic polarity* (i.e., positive or negative) of sentences/documents by analysis of the orientations of the individual words (Pang and Lee, 2004; Danescu-N.-M. et al., 2009; Kim and Hovy, 2004; Takamura et al., 2005). Sentiment dictionaries are utilized to facilitate the summarization. There are numerous works that, given a sentiment lexicon, analyze the structure of a sentence/document to infer its orientation, the holder of an opinion, the sentiment of the opinion, etc. (Breck et al., 2007; Ding and Liu, 2010; Kim and Hovy, 2004). Several domain independent sentiment dictionaries have been manually or (semi)-automatically created, e.g., General Inquirer (GI) (Stone et al., 1996), Opinion Finder (OF) (Wilson et al., 2005), Appraisal Lexicon (AL) (Taboada and Grieve, 2004), SentiWordNet (Baccianella et al., 2010) and Q-WordNet (Agerri and García-Serrano, 2010). Q-WordNet and SentiWordNet are lexical resources which classify the synsets(senses) in WordNet according to their polarities. We call them *sentiment sense dictionaries* (SSD). OF, GI and AL are called *sentiment word dictionaries* (SWD). They consist of words manually annotated with their corresponding polarities. The sentiment dictionaries have the following problems:

- They exhibit substantial (intra-dictionary) inaccuracies. For example, the synset
  {*Indo-European, Indo-Aryan, Aryan*} *(of or relating to the former Indo-European people)*,
  has a negative polarity in Q-WordNet, while most people would agree that this synset has a neutral polarity instead.
- They have (inter-dictionary) inconsistencies. For example, the adjective `cheap` is positive in AL and negative in OF.
- These dictionaries do not address the concept of polarity (in)consistency of words/synsets.

We concentrate on the concept of (in)consistency in this paper. We define consistency among the polarities of words/synsets in a dictionary and give methods to check it. A couple of examples help illustrate the problem we attempt to address.

The first example is the verbs `confute` and `disprove`, which have positive and negative polarities, respectively, in OF. According to WordNet, both words have a unique sense, which they share: *disprove, confute (prove to be false) "The physicist disproved his colleagues' theories"*
Assuming that WordNet has complete information about the two words, it is rather strange that the words have distinct polarities. By manually checking two other authoritative English dictionaries, Oxford[1] and Cambridge[2], we note that the information about `confute` and `disprove` in WordNet is the same as that in these dictionaries. So, the problem seems to originate in OF.

The second example is the verbs `tantalize` and `taunt`, which have positive and negative polarities, respectively, in OF. They also have a unique sense in WordNet, which they share. Again, there is a contradiction. In this case Oxford dictionary mentions a sense of `tantalize` that is missing from WordNet: *"excite the senses or desires of (someone)"*. This sense conveys a positive polarity. Hence, `tantalize` conveys a positive sentiment when used with this sense.

In summary, these dictionaries have conflicting information. Manual checking of sentiment dictionaries for inconsistency is a difficult endeavor. We deem words such as `confute` and `disprove` inconsistent. We aim to unearth these inconsistencies in sentiment dictionaries. The presence of inconsistencies found via polarity analysis is not exclusively attributed to one party, i.e., either the sentiment dictionary or WordNet. Instead, as emphasized by the above examples, some of them lie in the sentiment dictionaries, while others lie in WordNet. Therefore, a by-product of our polarity consistency analysis is that it can also locate some of the likely places where WordNet needs linguists' attention.

We show that the problem of checking whether the polarities of a set of words is consistent is NP-complete. Fortunately, the consistency problem can be reduced to the satisfiability problem (SAT). A fast SAT solver is utilized to detect inconsistencies and it is known such solvers can in practice determine consistency or detect inconsistencies. Experimental results show that substantial inconsistencies are discovered among words with polarities within and across sentiment dictionaries. This suggests that some remedial work needs to be performed on these sentiment dictionaries as well as on WordNet. The contributions of this paper are:

- address the consistency of polarities of words/senses. The problem has not been addressed before;
- show that the consistency problem is NP-complete;
- reduce the polarity consistency problem to the satisfiability problem and utilize a fast SAT solver to detect inconsistencies;
- give experimental results to demonstrate that our technique identifies considerable inconsistencies in various sentiment lexicons as well as discrepancies between these lexicons and WordNet.

2 Problem Definition

The polarities of the words in a sentiment dictionary may not necessarily be consistent (or correct). In this paper, we focus on the detection of polarity assignment inconsistencies for the words and synsets within and across dictionaries (e.g., OF vs. GI). We attempt to pinpoint the words with polarity inconsistencies and classify them (Section 3).

2.1 WordNet

We give a formal characterization of WordNet. This consists of words, synsets and frequency counts. A **word-synset network** $\mathcal{N}$ is quadruple $(\mathcal{W}, \mathcal{S}, \mathcal{E}, f)$ where $\mathcal{W}$ is a finite set of words, $\mathcal{S}$ is a finite set of synsets, $\mathcal{E}$ is a set of undirected edges between elements in $\mathcal{W}$ and $\mathcal{S}$, i.e., $\mathcal{E} \subseteq \mathcal{W} \times \mathcal{S}$ and f is a function assigning a positive integer to each element in $\mathcal{E}$. For an edge (w, s), $f(w, s)$ is called the **frequency of use** of w in the sense given by s. For any word w and synset s, we say that s is a synset of w if $(w, s) \in \mathcal{E}$. Also, for any word w, we let $freq(w)$ denote the sum of all $f(w, s)$ such that $(w, s) \in \mathcal{E}$. If a synset has a 0 frequency of use we replace it with 0.1, which is a standard smoothing technique (Han, 2005). For instance, the word `cheap` has four senses. The frequencies of occurrence of the word in the four senses are f_1 = 9, f_2 = 1, f_3 = 1 and f_4 = 0, respectively. By smoothing, $f_4 = 0.1$. Hence, $freq(\texttt{cheap}) = f_1 + f_2 + f_3 + f_4 = 11.1$. The **relative frequency** of the synset in the first sense of `cheap`, which denotes the probability that the word is used in the first sense, is $\frac{f_1}{freq(cheap)} = \frac{9}{11.1} = 0.81$.

[1] http://oxforddictionaries.com/
[2] http://dictionary.cambridge.org/

2.2 Consistent Polarity Assignment

We assume that each synset has a *unique* polarity. We define the polarity of a word to be a discrete probability distribution: P_+, P_-, P_0 with $P_+ + P_- + P_0 = 1$, where they represent the "likelihoods" that the word is positive, negative or neutral, respectively. We call this distribution a **polarity distribution**. For instance, the word `cheap` has the polarity distribution $P_+ = 0.81, P_- = 0.19$ and $P_0 = 0$. The polarity distribution of a word is estimated using the polarities of its underlying synsets. For instance `cheap` has four senses, with the first sense being positive and the last three senses being negative. The probability that the word expresses a negative sentiment is $P_- = \frac{f_2+f_3+f_4}{freq(cheap)} = 0.19$, while the probability that the word expresses a positive sentiment is $P_+ = \frac{f_1}{freq(cheap)} = 0.81$. $P_0 = 1 - P_+ - P_- = 0$.

Our view of characterizing the polarity of a word using a polarity distribution is shared with other previous works (Kim and Hovy, 2006; Andreevskaia and Bergler, 2006). Nonetheless, we depart from these works in the following key aspect. We say that a word has a (mostly) positive (negative) polarity if the **majority sense of the word** is positive (negative). That is, a word has a mostly positive polarity if $P_+ > P_- + P_0$ and it has a mostly negative polarity if $P_- > P_+ + P_0$. Or, equivalently, if $P+ > \frac{1}{2}$ or $P_- > \frac{1}{2}$, respectively. For example, on majority, `cheap` conveys positive polarity since $P_+ = .081 > \frac{1}{2}$, i.e., the majority sense of the word `cheap` has positive connotation. We conducted empirical studies using Micro-WN(Op)[3] and GI, and empirically showed that the majority sense property is the underlying property of domain independent SWD and captures the "collective behavior" of human annotators (TechRep, 2012).

Based on this study, we contend that GI, OF and AL tacitly assume this property. For example, the verb `steal` is assigned only negative polarity in GI. This word has two other less frequently occurring senses, which have positive polarities. The polarity of `steal` according to these two senses is not mentioned in GI. This is the case for the overwhelming majority of the entries in the three dictionaries: only 112 out of a total of 14,105 entries in the three dictionaries regard words with multiple polarities. For example, the verb `arrest` is mentioned with both negative and positive polarities in GI. We regard an entry in an SWD as the majority sense of the word has the specified polarity, although the word may carry other polarities. For instance, the adjective `cheap` has positive polarity in GI. The only assumption we make about the word is that it has a polarity distribution such that $P_+ > P_- + P_0$. This interpretation is consistent with the senses of the word. In this work we show that this property allows the polarities of words in input sentiment dictionaries to be checked. We formally state this property.

Definition 1. *Let w be a word and S_w its set of synsets. Each synset in S_w has an associated polarity and a relative frequency with respect to w. w has polarity p, $p \in \{$positive, negative$\}$ if there is a subset of synsets $S' \subseteq S_w$ such that each synset $s \in S'$ has polarity p and $\sum_{s \in S'} \frac{f(w,s)}{freq(w)} > 0.5$. S' is called a **polarity dominant subset**. If there is no such subset then w has a neutral polarity.*

*$S' \subseteq S_w$ is a **minimally dominant subset of synsets** (MDSs) if the sum of the relative frequencies of the synsets in S' is larger than 0.5 and the removal of any synset s from S' will make the sum of the relative frequencies of the synsets in $S' - \{s\}$ smaller than or equal to 0.5.*

The definition does not preclude a word from having a polarity with a *majority* sense and a different polarity with a *minority* sense. For example, the definition does not prevent a word from having both positive and negative senses, but it prevents a word from concomitantly having a majority sense of being positive and a majority sense of being negative.

Despite using a "hard-coded" constant in the definition, our approach is generic and does not dependent on the constant 0.5. This constant is just a lower bound for deciding whether a word has a majority sense with a certain polarity. It also is intuitively appealing. The constant can be replaced with an arbitrary threshold τ between 0.5 and 1.

We need a formal description of polarity assignments to the words and synsets in WordNet. We assign polarities from the set $\mathcal{P} = \{$positive, negative, neutral$\}$ to elements in $\mathcal{W} \cup \mathcal{S}$. Formally, a polarity assignment γ for a network $\mathcal{N}$ is a function from $\mathcal{W} \cup \mathcal{S}$ to the set $\mathcal{P}$. Let γ be a polarity assignment for $\mathcal{N}$. We say that γ is **consistent** if it satisfies the following condition for each $w \in \mathcal{W}$:

For $p \in \{$positive, negative$\}$, $\gamma(w) = p$ iff the

[3] http://www-3.unipv.it/wnop/

Table 1: Disagreement between dictionaries.

Pairs of Dictionaries	Word Polarity Disagreement	
	Inconsistency	Overlap
OF & GI	90	2,924
OF & AL	73	1,150
GI & AL	18	712

sum of all $f(w,s)$ such that $(w,s) \in \mathcal{E}$ and $\gamma(s) = p$, is greater than $\frac{freq(w)}{2}$. Note that, for any $w \in \mathcal{W}$, $\gamma(w) = neutral$ iff the above inequality is not satisfied for both values of p in {positive, negative}.

We contend that our approach is applicable to domain dependent sentiment dictionaries, too. We can employ WordNet Domains (Bentivogli et al., 2004). WordNet Domains augments WordNet with domain labels. Hence, we can project the words/synsets in WordNet according to a domain label and then apply our methodology to the projection.

3 Inconsistency Classification

Polarity inconsistencies are of two types: input and complex. We discuss them in this section.

3.1 Input Dictionaries Polarity Inconsistency

Input polarity inconsistencies are of two types: intra-dictionary and inter-dictionary inconsistencies. The latter are obtained by comparing (1) two SWDs, (2) an SWD with an SSD and (3) two SSDs.

3.1.1 Intra-dictionary inconsistency

An SWD may have triplets of the form (w, pos, p) and (w, pos, p'), where $p \neq p'$. For instance, the verb `brag` has both positive and negative polarities in OF. For these cases, we look up WordNet and apply Definition 1 to determine the polarity of word w with part of speech pos. The verb `brag` has negative polarity according to Definition 1. Such cases simply say that the team who constructs the dictionary believes the word has multiple polarities as they do not adopt our dominant sense principle. There are 58 occurrences of this type of inconsistency in GI, OF and AL. Q-WordNet, a sentiment sense dictionary, does not have intra-inconsistencies as it does do not have a synset with multiple polarities.

3.1.2 Inter-dictionary inconsistency

A word belongs to this category if it appears with different polarities in different SWDs. For instance, the adjective `joyless` has positive polarity in OF and negative polarity in GI. Table 1 depicts the overlapping relationships between the three SWDs: e.g., OF has 2,933 words in common with GI. The three dictionaries largely agree on the polarities of the words they pairwise share. For instance, out of 2,924 words shared by OF and GI, 2,834 have the same polarities. However, there are also a significant number of words which have different polarities across dictionaries. Case in point, OF and GI disagree on the polarities of 90 words. Among the three dictionaries there are 181 polarity inconsistent words. These words are manually corrected using Definition 1 before the polarity consistency checking is applied to the union of the three dictionaries. This union is called *disagreement-free union*.

3.2 Complex Polarity Inconsistency

This kind of inconsistency is more subtle and cannot be detected by direct comparison of words/synsets. They consist of *sets* of words and/or synsets whose polarities cannot concomitantly be satisfied. Recall the example of the verbs `confute` and `disprove` in OF given in Section 1. Recall our argument that by assuming that WordNet is correct, it is not possible for the two words to have different polarities: the sole synset, which they share, would have two different polarities, which is a contradiction.

The occurrence of an inconsistency points out the presence of incorrect input data:

- the information given in WordNet is incorrect, or
- the information in the given sentiment dictionary is incorrect, or both.

Regarding WordNet, the errors may be due to (1) a word has senses that are missing from WordNet or (2) the frequency count of a synset is inaccurate. A comprehensive analysis of every synset/word with inconsistency is a tantalizing endeavor requiring not only a careful study of multiple sources (e.g., dictionaries such as Oxford and Cambridge) but also linguistic expertise. It is beyond the scope of this paper to enlist all potentially inconsistent words/synsets and the possible remedies. Instead, we limit ourselves to drawing attention to the occurrence of these issues through examples, welcoming experts in the area to join the corrective efforts. We give more examples of inconsistencies in order to illustrate additional discrepancies between input dictionaries.

3.2.1 WordNet vs. Sentiment Dictionaries

The adjective `bully` is an example of a discrepancy between WordNet and a sentiment dictionary. The word has negative polarity in OF and has a sin-

gle sense in WordNet. The sense is shared with the word `nifty`, which has positive polarity in OF. By applying Definition 1 to `nifty` we obtain that the sense is positive, which in turn, by Definition 1, implies that `bully` is positive. This contradicts the input polarity of `bully`. According to the Webster dictionary, the word has a sense (i.e., *resembling or characteristic of a bully*) which has a negative polarity, but it is not present in WordNet. The example shows the presence of a discrepancy between WordNet and OF, namely, OF seems to assign polarity to a word according to a sense that is not in WordNet.

3.2.2 Across Sentiment Dictionaries

We provide examples of inconsistencies across sentiment dictionaries here. Our first example is obtained by comparing SWDs. The adjective `comic` has negative polarity in AL and the adjective `laughable` has positive polarity in OF. Through deduction (i.e., by successive applications of Definition 1), the word `risible`, which is not present in either of the dictionaries, is assigned negative polarity because of `comic` and is assigned positive polarity because of `laughable`.

The second example illustrates that an SWD and an SSD may have contradicting information. The verb `intoxicate` has three synsets in WordNet, each with the same frequency. Hence, their relative frequencies with respect to `intoxicate` are $\frac{1}{3}$. On one hand, `intoxicate` has a negative polarity in GI. This means that $P_- > \frac{1}{2}$. On the other hand, two of its three synsets have positive polarity in Q-WordNet. So, $P_+ = \frac{2}{3} > \frac{1}{2}$, which means that $P_- < \frac{1}{2}$. This is a contradiction.

3.3 Consistent Polarity Assignment

Given the discussion above, it clearly is important to find all occurrences of inconsistent words. This in turn boils down to finding those words with the property that there does not exist any polarity assignment to the synsets, which is consistent with their polarities. It turns out that the complexity of the problem of assigning polarities to the synsets such that the assignment is consistent with the polarities of the input words, called `Consistent Polarity Assignment` problem, is a "hard" problem, as described below. The problem is stated as follows:

Consider two sets of nodes of type synsets and type words, in which each synset of a word has a relative frequency with respect to the word. Each synset can be assigned a positive, negative or neutral polarity. A word has polarity p if it satisfies the hypothesis of Definition 1. The question to be answered is: Given an assignment of polarities to the words, does there exist an assignment of polarities to the synsets that agrees with that of the words?

In other words, given the polarities of a subset of words (e.g., that given by one of the three SWDs) the problem of finding the polarities of the synsets that agree with this assignment is a "hard" problem.

Theorem 1. *The Consistent Polarity Assignment problem is NP-complete.*

4 Polarity Consistency Checking

To "exhaustively" solve the problem of finding the polarity inconsistencies in an SWD, we propose a solution that reduces an instance of the problem to an instance of CNF-SAT. We can then employ a fast SAT solver (e.g., (Xu et al., 2008; Babic et al., 2006)) to solve our problem. CNF-SAT is a decision problem of determining if there is an assignment of True and False to the variables of a Boolean formula Φ in conjunctive normal form (CNF) such that Φ evaluates to True. A formula is in CNF if it is a conjunction of one or more clauses, each of which is a disjunction of literals. CNF-SAT is a classic NP-complete problem, but, modern SAT solvers are capable of solving many practical instances of the problem. Since, in general, there is no easy way to tell the difficulty of a problem without trying it, SAT solvers include time-outs, so they will terminate even if they cannot find a solution.

We developed a method of converting an instance of the polarity consistency checking problem into an instance of CNF-SAT, which we will describe next.

4.1 Conversion to CNF-SAT

The input consists of an SWD D and the word-synset network $\mathcal{N}$. We partition $\mathcal{N}$ into connected components. For each synset s we define three Boolean variables s_-, s_+ and s_0, corresponding to the negative, positive and neutral polarities, respectively. In this section we use $-, +, 0$ to denote negative, positive and neutral polarities, respectively.

Let Φ be the Boolean formula for a connected component M of the word-synset network $\mathcal{N}$. We introduce its clauses. First, for each synset s we need a clause $C(s)$ that expresses that the synset can have

only one of the three polarities: $C(s) = (s_+ \wedge \neg s_- \wedge \neg s_0) \vee (s_- \wedge \neg s_+ \wedge \neg s_0) \vee (s_0 \wedge \neg s_- \wedge \neg s_+)$.

Since a word has a neutral polarity if it has neither positive nor negative polarities, we have that $s_0 = \neg s_+ \wedge \neg s_-$. Replacing this expression in the equation above and applying standard Boolean logic formulas, we can reduce it to

$$C(s) = \neg s_+ \vee \neg s_- \quad (1)$$

For each word w with polarity $p \in \{-, +, 0\}$ in D we need a clause $C(w, p)$ that states that w has polarity p. So, the Boolean formula for a connected component M of the word-synset network $\mathcal{N}$ is:

$$\Phi = \bigwedge_{s \in M} C(s) \wedge \bigwedge_{(w,p) \in D} C(w, p). \quad (2)$$

From Definition 1, w is neutral if it is neither positive nor negative. Hence, $C(w, 0) = \neg C(w, -) \wedge \neg C(w, +)$. So, we need to define only the clauses $C(w, -)$ and $C(w, +)$, which correspond to w having polarity negative and positive, respectively. So, herein $p \in \{-, +\}$, unless otherwise specified.

Our method is based on the following statement in Definition 1: w has polarity p if there exists a polarity dominant subset among its synsets. Thus, $C(w, p)$ is defined by enumerating all the MDSs of w. If at least one of them is a polarity dominant subset then $C(w, p)$ evaluates to True.

Exhaustive Enumeration of MDSs Method (EEM) We now elaborate the construction of $C(w, p)$. We enumerate all the MDSs of w and for each of them we introduce a clause. The clauses are then concatenated by OR in the Boolean formula. Let $C(w, p, T)$ denote the clause for an MDS T of w, when w has polarity $p \in \{-, +\}$. Hence,

$$C(w, p) = \bigvee_{T \in MDS(w)} C(w, p, T), \quad (3)$$

where $MDS(w)$ is the set of all MDSs of w.

For each MDS T of w, the clause $C(w, p, T)$ is the AND of the variables corresponding to polarity p of the synsets in T. That is,

$$C(w, p, T) = \bigwedge_{s \in T} s_p,\ p \in \{-, +\}. \quad (4)$$

The formula Φ is not in CNF after this construction and it needs to be converted. The conversion to CNF is a standard procedure and we omit it in this paper. Φ in CNF is input to a SAT solver.

Example 1. *Consider a connected component consisting of the words w = `cheap`, v = `inexpensive` and u = `sleazy`. `cheap` has a positive polarity, whereas `inexpensive` and `sleazy` have negative polarities. The synsets of these words are: $\{s^1, s^2, s^3, s^4\}$, $\{s^1\}$ and $\{s^3, s^4, s^5\}$, respectively (refer to WordNet). The relative frequencies of s^3, s^4 and s^5 w.r.t. `sleazy` are all equal to 1/3. We have 15 binary variables, 3 per synset, $s^i_-, s^i_+, s^i_0, 1 \leq i \leq 5$. The only MDS of `cheap` is $\{s^1\}$, which coincides with that of `inexpensive`. Those of `sleazy` are $\{s^3, s^4\}$, $\{s^3, s^5\}$ and $\{s^4, s^5\}$. For each s^i we need a clause $C(s^i)$. Hence, $C(w, +) = s^1_+$, $C(v, -) = s^1_-$ and $C(u, -) = (s^3_- \wedge s^4_-) \vee (s^3_- \wedge s^5_-) \vee (s^4_- \wedge s^5_-)$. Thus, $\Phi = \bigwedge_i C(s^i) \wedge [s^1_+ \wedge s^1_- \wedge ((s^3_- \wedge s^4_-) \vee (s^3_- \wedge s^5_-) \vee (s^4_- \wedge s^5_-))]$. Φ is not in CNF and needs to be converted. For Φ to be True, the clauses $C(w, +) = s^1_+$ and $C(v, -) = s^1_-$ must be True. But, this makes $C(s^1)$ False. Hence, Φ is not satisfiable. The clauses $C(w, +) = s^1_+$ and $C(v, -) = s^1_-$ are unsatisfiable and thus the polarities of `cheap` and `inexpensive` are inconsistent.*

4.2 Implementation Issues

The above reduction is exponential in the number of clauses (see, Equation 3) in the worst case. A polynomial reduction is possible, but it is significantly more complicated to implement. We choose to present the exponential reduction in this paper because it can handle over 97% of the words in WordNet and it is better suited to explain one of the main contributions of paper: the translation from the polarity consistency problem to SAT.

WordNet possesses nice properties, which allows the exponential reduction to run efficiently in practice. First, 97.2% of its (word, part-of-speech) pairs have 4 or fewer synsets. Thus, these words add very few clauses to a CNF formula (Equation 3). Second, WordNet can be partitioned into 33,015 *non-trivial* connected components, each of which corresponds to a Boolean formula and they all are independently handled. A non-trivial connected component has at least two words. Finally, in practice, not all connected components need to be considered for an input sentiment dictionary D, but only those having at least two words in D. In our experiments the largest number of components that need to be processed is

Table 2: Distribution of words and synsets

POS	Words	Synsets	OF	GI	AL	QWN
Noun	117,798	82,115	1,907	1,444	2	7,403
Verb	11,529	13,767	1,501	1,041	0	4006
Adjective	21,479	18,156	2,608	1,188	1,440	4050
Adverb	4,481	3,621	775	51	317	40
Total	**155,287**	**117,659**	**6,791**	**3,961**	**1,759**	**15,510**

1,581, for the disagreement-free union dictionary.

5 Detecting Inconsistencies

In this section we describe how we detect the words with polarity inconsistencies using the output of a SAT solver. For an unsatisfiable formula, a modern SAT solver returns a *minimal unsatisfiable core* (MUC) from the original formula. An *unsatisfiable core* is minimal if it becomes satisfiable whenever any one of its clauses is removed. There are no known practical algorithms for computing the *minimum core* (Dershowitz et al., 2006). In our problem a MUC corresponds to a set of polarity inconsistent words. The argument is as follows. Consider W the set of words in a connected component and Φ the CNF formula generated with the above method. During the transformation we keep track of the clauses introduced in Φ by each word. Suppose Φ is inconsistent. Then, the SAT solver returns a MUC. Each clause in a MUC is mapped back to its corresponding word(s). We obtain the corresponding subset of words $W', W' \subseteq W$. Suppose that Φ' is the Boolean CNF formula for the words in W'. The set of clauses in Φ' is a subset of those in Φ. Also, the clauses in the MUC appear in Φ'. Thus, Φ' is unsatisfiable and the words in W' are inconsistent.

To find *all* inconsistent words we ought to generate all MUCs. Unfortunately, this is a "hard" problem (Dershowitz et al., 2006) and no open source SAT solver possesses this functionality. We however observe that the two SAT solvers we use for our experiments (SAT4j and PicoSAT (Biere, 2008)) return different MUCs for the same formula and we use them to find as many inconsistencies as possible.

6 Experiments

The goal of the experimental study is to show that our techniques can identify considerable inconsistencies in various sentiment dictionaries.

Table 3: Intra- and inter-dictionaries inconsistency

POS	OF	QW	GI	QW	AL	QW	UF	QW
Noun	23	119	4	61	0	42	90	140
Verb	66	113	2	67	0	0	63	137
Adj.	90	170	8	48	0	0	27	177
Adv.	61	1	0	0	2	0	69	1
Total	**240**	**403**	**14**	**176**	**2**	**42**	**249**	**455**

Data sets In our experiments, we use WordNet 3.0, GI, OF, AL and Q-WordNet. Their statistics are given in Table 2. The table shows the distribution of the words and synsets per part of speech. Columns 2 and 3 pertain to WordNet. There are 3,961 entries in `GI`, 1,759 entries in `AL` and 6,791 entries in `OF` which appear in WordNet. Q-WordNet has 15,510 entries, i.e., synsets with polarities.

Inconsistency Detection We applied our method to (1) each of AL, GI and OF; (2) the disagreement-free union (UF); (3) each of AL, GI and OF together with Q-WordNet and (4) UF and Q-WordNet. Table 3 summarizes the outcome of the experimental study. EEM finds 240, 14 and 2 polarity inconsistent words in OF, GI and AL, respectively. The ratio between the number of inconsistent words and the number of input words is the highest for OF and the lowest for AL. The union dictionary has 7,794 words and 249 out of them are found to be polarity inconsistent words. Recall that we manually corrected the polarities of 181 words, to the best of our understanding. So, in effect the three dictionaries have 249 + 181 = 430 polarity inconsistent words. As discussed in the previous section, these may not be all the polarity inconsistencies in UF. In general, to find all inconsistencies we need to generate all MUCs. Generating all MUCs is an "overkill" and the SAT solvers we use do not implement such a functionality. In addition, the intention of SAT solver designers is to use MUCs in a interactive manner. That is, the errors pointed out by a MUC are corrected and then the new improved formula is re-evaluated by the SAT solver. If an error is still present a new MUC is reported, and the process repeats until the formula has no errors. Or, in our problem, until a dictionary is consistent.

We also paired Q-WordNet with each of the SWDs. Table 3 presents the results. Observe that polarities assigned to the words in AL and GI largely agree with the polarities assigned to the synsets in

Q-WordNet. This is expected for AL because it has only two nouns and no verb, while Q-WordNet has only 40 adverbs. Consequently, these two dictionaries have limited "overlay". The union dictionary and Q-WordNet have substantial inconsistencies: the polarity of 455 words in the union dictionary disagrees with the polarities assigned to their underlying synsets in Q-WordNet.

Sentence Level Evaluation We took 10 pairs of inconsistent words per part of speech; in total, we collected a set IW of 80 inconsistent words. Let $\langle w, pos, p\rangle \in IW$, p is the polarity of w. We collected 5 sentences for $\langle w, pos\rangle$ from the set of snippets returned by Google for query w. We parsed the snippets and identified the first 5 occurrences of w with the part of speech pos. Then two graduate students with English background analyzed the polarities of $\langle w, pos\rangle$ in the 5 sentences. We counted the number of times $\langle w, pos\rangle$ appears with polarity p and polarities different from p. We defined an agreement scale: total agreement (5/5), most agreement (4/5), majority agreement (3/5), majority disagreement (2/5), most disagreement (1/5), total disagreement (0/5). We computed the percentage of words per agreement category. We repeated the experiment for 40 randomly drawn words (10 per part of speech) from the set of consistent words. In total 600 sentences were manually analyzed. Figure 1 shows the distribution of the (in)consistent words. For example, the annotators totally agree with the polarities of 55% of the consistent words, whereas they only totally agree with 16% of the polarities of the inconsistent words. The graph suggests that the annotators disagree to some extent (total disagreement + most disagreement + major disagreement) with 40% of the polarities of the inconsistent words, whereas they disagree to some extent with only 5% of the consistent words. We also manually investigated the senses of these words in WordNet. We noted that 36 of the 80 inconsistent words (45%) have missing senses according to one of these English dictionaries: Oxford and Cambridge.

Computational Issues We used a 4-core CPU computer with 12GB of memory. EEM requires 10GB of memory and cannot handle words with more than 200,000 MDSs: for UF we left the SAT solver running for a week without ever terminating. In contrast, it takes about 4 hours if we limit the set of words to those that have up to 200,000 MDSs. EEM could not handle words such as `make`, `give` and `break`. Recall however that we did not generate all MUCs. We do not know how long would that might have taken. (The polynomial method handles all the words in WordNet and it takes 5GB of memory and about 2 hours to finish.)

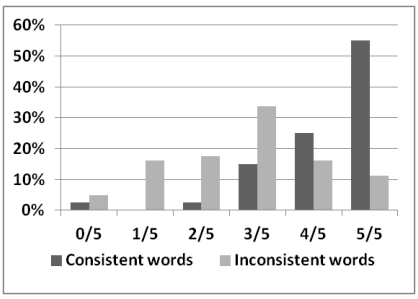

Figure 1: Human classification of (in)consistent words.

7 Related Work

Several researchers have studied the problem of finding opinion words (Liu, 2010). There are two lines of work on sentiment polarity lexicon induction: corpora-based (Hatzivassiloglou and McKeown, 1997; Kanayama and Nasukawa, 2006; Qiu et al., 2009; Wiebe, 2000) and dictionary-based (Andreevskaia and Bergler, 2006; Agerri and García-Serrano, 2010; Dragut et al., 2010; Esuli and Sebastiani, 2005; Baccianella et al., 2010; Hu and Liu, 2004; Kamps et al., 2004; Kim and Hovy, 2006; Rao and Ravichandran, 2009; Takamura et al., 2005). Our work falls into the latter. Most of these works use the lexical relations defined in WordNet (e.g., synonym, antonym) to derive sentiment lexicons. To our knowledge, none of the earlier works studied the problem of polarity consistency checking for a sentiment dictionary. Our techniques can pinpoint the inconsistencies within individual dictionaries and across dictionaries.

8 Conclusion

We studied the problem of checking polarity consistency for sentiment word dictionaries. We proved that this problem is NP-complete. We showed that in practice polarity inconsistencies of words both within a dictionary and across dictionaries can be obtained using an SAT solver. The inconsistencies are pinpointed and this allows the dictionaries to be improved. We reported experiments on four sentiment dictionaries and their union dictionary.

References

Rodrigo Agerri and Ana García-Serrano. 2010. Q-wordnet: Extracting polarity from wordnet senses. In *LREC*.

A. Andreevskaia and S. Bergler. 2006. Mining wordnet for fuzzy sentiment: Sentiment tag extraction from wordnet glosses. In *EACL*.

Domagoj Babic, Jesse Bingham, and Alan J. Hu. 2006. B-cubing: New possibilities for efficient sat-solving. *TC*, 55(11).

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May.

Luisa Bentivogli, Pamela Forner, Bernardo Magnini, and Emanuele Pianta. 2004. Revising the wordnet domains hierarchy: semantics, coverage and balancing. MLR.

Armin Biere. 2008. PicoSAT essentials. *JSAT*, 4(2-4):75–97.

Eric Breck, Yejin Choi, and Claire Cardie. 2007. Identifying expressions of opinion in context. In *IJCAI*.

Cristian Danescu-N.-M., Gueorgi Kossinets, Jon Kleinberg, and Lillian Lee. 2009. How opinions are received by online communities: a case study on amazon.com helpfulness votes. In *WWW*, pages 141–150.

Nachum Dershowitz, Ziyad Hanna, and Er Nadel. 2006. A scalable algorithm for minimal unsatisfiable core extraction. In *In Proc. SAT06*. Springer.

Xiaowen Ding and Bing Liu. 2010. Resolving object and attribute coreference in opinion mining. In *COLING*.

Eduard C. Dragut, Clement T. Yu, A. Prasad Sistla, and Weiyi Meng. 2010. Construction of a sentimental word dictionary. In *CIKM*, pages 1761–1764.

Andrea Esuli and Fabrizio Sebastiani. 2005. Determining the semantic orientation of terms through gloss classification. In *CIKM*, pages 617–624.

A. Esuli and F. Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *LREC*.

Murthy Ganapathibhotla and Bing Liu. 2008. Mining opinions in comparative sentences. In *COLING*.

Michael R. Garey and David S. Johnson. 1990. *Computers and Intractability; A Guide to the Theory of NP-Completeness*.

Jiawei Han. 2005. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, ACL '98, pages 174–181, Stroudsburg, PA, USA. Association for Computational Linguistics.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

J. Kamps, M. Marx, R. Mokken, and M. de Rijke. 2004. Using wordnet to measure semantic orientation of adjectives. In *LREC*.

Hiroshi Kanayama and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, pages 355–363, Stroudsburg, PA, USA. Association for Computational Linguistics.

M. Kim and E. Hovy. 2004. Determining the sentiment of opinions. In *COLING*.

Soo-Min Kim and Eduard Hovy. 2006. Identifying and analyzing judgment opinions. In *HLT-NAACL*.

Bing Liu. 2010. Sentiment analysis and subjectivity. In Nitin Indurkhya and Fred J. Damerau, editors, *Handbook of Natural Language Processing, Second Edition*. CRC Press, Taylor and Francis Group, Boca Raton, FL. ISBN 978-1420085921.

J. Martin. 2000. Beyond exchange: Appraisal systems in english. In *Evaluation in Text. Oxford: Oxford University Press.*, pages 142–175.

B. Pang and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *ACL*.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2009. Expanding domain sentiment lexicon through double propagation. In *Proceedings of the 21st international jont conference on Artifical intelligence*, IJCAI'09, pages 1199–1204.

Delip Rao and Deepak Ravichandran. 2009. Semi-supervised polarity lexicon induction. In *EACL*.

P. Stone, D. Dunphy, M. Smith, and J. Ogilvie. 1996. The general inquirer: A computer approach to content analysis. In *MIT Press*.

M. Taboada and J. Grieve. 2004. Analyzing appraisal automatically. In *AAAI Spring Symposium*.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *ACL*, pages 133–140.

TechRep. 2012. Technical report: Polarity consistency checking for sentiment dictionaries.

Janyce Wiebe. 2000. Learning subjective adjectives from corpora. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and*

Twelfth Conference on Innovative Applications of Artificial Intelligence, pages 735–740. AAAI Press.

Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. 2004. Just how mad are you? finding strong and weak opinion clauses. In *AAAI'04*, pages 761–767.

T. Wilson, J. Wiebe, and P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT/EMNLP*.

Lin Xu, Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. 2008. Satzilla: portfolio-based algorithm selection for sat. *J. Artif. Int. Res.*, 32:565–606, June.