SenseNet: A Linguistic Tool to Visualize Numerical-Valance Based Sentiment of Textual Data

Shaikh Mostafa Al Masum, Helmut Prendinger[1] and Ishizuka Mitsuru

Department of Information and Communication Engineering
University of Tokyo,
7-3-1 Hongo, Bunkyo-Ku, 113-8656 Tokyo, Japan
[1]Digital Contents and Media,
Sciences Research Division,
National Institute of Informatics,
2-1-2 Hitotsubashi, Chiyoda Ku, 101-8430 Tokyo, Japan
E-mail: almasum@gmail.com, helmut@nii.ac.jp, ishizuka@ieee.org

Abstract

We believe that text is an important modality for human-computer interaction; so studying the relationship between sentiments conveyed through text and natural language by a numerical analysis strategy is worthwhile for the task of sentiment analysis. Different approaches have already been employed to "sense" sentiment, especially from the texts, but none of those ever considered the valenced based cognitive and appraisal structure of sentiments. The paper puts forward a different approach to sense and visualize sentiments embedded in texts by applying a numerical-valence based analysis. To meet this objective we have developed a linguistic tool called SenseNet to visualize sentiments of input texts.

1 Introduction

The assessment of sentiment in written text is inevitably subjective and subject to considerable disagreement (Wiebe et al., 2001) but the interest in sentiment based automated text categorization has increased with the availability of large amount of text on the Internet. The applications range from document organization, automatic document indexing for information retrieval, text or email filtering, word sense disambiguation, categorization of web pages, news-article classification and, most recently, spam filters. It is noticed that all the previous approaches for analyzing texts for sentiment have commonly employed techniques like, keyword spotting (Boucouvalas and Zhe, 2002); lexical affinity (Valitutti et al., 2004); statistical methods (Pennebaker et al., 2001); pre-processed models (Elliott, 1994); a dictionary of affective concepts and lexicon (Rosis and Grasso, 2000); commonsense knowledgebase (Liu et al., 2003); naive Bayesian method (Sebastiani, 2002); and support vector machines (Maria and Silva, 2001), but none of those ever considered the valenced based cognitive and appraisal structure of sentiments. Approach mentioned by (Nasukawa and Yi, 2003) used a sentiment analysis dictionary having 3,513 entries and instead of analyzing the favorability of the whole context each statement on favorability is extracted, and present them to the end users so that they can use the results according to their application requirements. But the system outputs -1 to indicate a negative sentiment (due to shallow understanding) for the sentence "*It's difficult to take a bad picture with this camera.*", whereas this is a positive statement for the camera. According to a linguistic survey (Pennebaker *et al.*, 2001) across all of the studies described, only 4% were emotional (adjectives) words used in written texts. This indicates that affective lexicons might not be sufficient to recognize affective information from texts and raises the suspicion that a machine learning or keyword spotting method might not perform well for this objective. The main idea of SenseNet is to create a graph of lexical-units from the input sentence(s); assigns a numerical value to those based on their sense affinity; assesses the values using rules; and finally outputs sense-valence for each lexical-unit (e.g. sentence, paragraph) and a special browser then visualizes sentiments by different symbols.

2 SenseNet Architecture

We admit that analysis of favorable or unfavorable opinions is a task requiring emotional intelligence and deep understanding of the textual context, involving common-sense and domain knowledge as well as linguistic knowledge. The interpretation of opinions is usually debatable affair even for humans. However our approach is an attempt towards this task. SenseNet consists of WordNet 2.1; ConceptNet 2.1; a Knowledge-base; a set of Assumptions; a Language Parser and Sense-browser that displays sentiment of each line of the input text-chunk in terms of numerical valence and icons.

2.1 Language Parser

We are using the *Machinese Syntax* (Connexor Oy) program to obtain XML-formatted shallow-parsed information for an input sentence for further processing. For example, for the input sentence, "*Two members of Tonga's royal family were killed when a teenager racing her car crashed into their vehicle*", we obtain XML-like syntactical information from the parser, which is further processed to output as tuples of Subject, Subject Type, Subject Attributes; Action, Action Status, Action Attribute; Object, Object Type and Object Attribute, as indicated in Figure 1.

```
[[['Subject:' 'member', 'Subject Type:' 'Person', 'Subject
Attrib:' ['quantity: two', 'N GEN SG: tonga', 'A ABS:
royal', 'N NOM: family']]
['Action:' 'kill', 'Action Status:' 'Past Particle', 'Action
Attrib: ['time: when']],
['Concept:' '', 'Concept Type:' '', 'Concept Attrib:'']]],
[['Subject: ' 'teenager', 'Subject Type:' 'Person', 'Subject
Attrib: [ ]],
['Action:' 'race', 'Action Status:' 'Continuous', 'Action
Attrib:'[]],
['Concept:' 'car', 'Concept Type:' 'N NOM', 'Concept
Attrib:' ['PRON PERS GEN SG3:she']]],
[['Subject:' 'car', 'Subject Type:' 'Other 3rd', 'Subject
Attrib:' [ ]],
['Action:' 'crash', 'Action Status:' 'past', 'Action Attrib:'
['goal: vehicle']],
['Concept:' 'vehicle', 'Concept Type:' 'N NOM', 'Concept
Attrib:' ['PRON PERS GEN PL3: they']]]
```

Figure 1 Output of deep-parse

A tuple encodes information about "*who is associated with what and how*". The output given in Figure 1 has three such tuples.

2.2 WordNet

The WordNet (Fellbaum, 1999) is a database of English words organized into synonym sets, and each word is linked by a small set of semantic relations such as the synonym relation and 'is-a' hierarchical relations representing one underlying lexical concept. The current version 2.1 contains 207016 word-sense pairs and 78695 polysemous senses. The SenseNet is employing WordNet 2.1 for two purposes. The primary purpose is to assign a numerical value (either positive or negative) to each of our enlisted words based on manual investigation of senses of each word done by a group of students and volunteers (explained in section 2.4). The secondary purpose is to obtain the synonyms for a word that is not found in the SenseNet list and to scrutinize this list with respect to built in list for which numerical values are assigned.

2.3 ConceptNet

ConceptNet (Liu and Singh, 2004) is a semantic network of common-sense knowledge that at present contains 1.6 million edges connecting more than 300000 nodes. Nodes are semi-structured English fragments, interrelated by ontology of twenty semantic relations encompassing the spatial, physical, social, temporal, and psychological aspects of everyday life. ConceptNet is generated automatically from the 700000 sentences of the Open Mind Common Sense (OMCS) Project which is World Wide Web based collaboration with over 14000 authors. In the SenseNet we have employed ConceptNet to utilize *DisplayNode()* function that returns all the possible semantic relationships of an input concept. How we processed the output of this function is explained in the section 2.5.

2.4 Knowledge-base

SenseNet maintains a list of scored verbs, adjectives, adverbs, concepts and named-entities. These lists are utilized with the rules to score the sentiment valence of each sentence. The motivation of scoring the words is to know the general lexical-affinity of a word to a particular sentiment, for example, *'destroy'* usually indicates negative connotation but *'develop'* gives positive.

Scored-List of Action, Adjective and Adverb: A group of students and volunteers have manually counted the positive and negative senses of each word of our customized list of verbs and adjectives according to the contextual understanding of each sense appeared in WordNet 2.1; and thus we maintain a database of scored verbs, adjectives and adverbs. An individual's score of a verb is stored as following tuple-like format; verb-word [*Positive-Sense Count, Negative- Sense Count, Prospective Value, Praiseworthy Value, Polarity*

Value]. For example, for the word *'attack'* WordNet 2.1 outputs 6 senses as a verb and someone may consider 5 senses as negative and 1 sense as positive whereas another may consider 3 negative and 3 positive senses. Thus we collected such counts for our listed words and the following formulae (scale of -5 to 5) are used to assign the polarity, prospective and praiseworthy values to each action word. Only polarity values are calculated for adjectives.

- Polarity Value = Average (((Positive-Sense Count – Negative-Sense Count)/Total Sense Count) * 5.0) …………………………….(1)
- Prospective Value= Average ((Positive-Sense Count / Total Sense Count) * 5.0)...(2)
- Praiseworthy Value = Average (Polarity Value + Prospective Value) ..………… (3)

At present we scored 723 verbs, 205 phrasal verbs, 237 adjectives related to shape, time, sound, taste/touch, condition, appearance and 711 adjectives related to emotional affinity and 144 adverbs.

Scored-List of Abstract-Concept: By the term 'Abstract-Concept' we mean a group of concepts known as 'key-concepts' that are described by a set of descriptive properties and contain a valence value (negative, positive or neutral) called as 'concept-valence' and may belong to a set of emotion-type. SenseNet maintains a database of such 'Abstract-Concept'. For example, we have *'Religion'* as an abstract-concept which describes about the group of concepts pertaining to religious affairs and thus *'Religion'* inducts the key-concepts like 'god', 'heaven', 'church', 'prayer', 'faith', 'islam', 'christianity', 'judaism', 'hell', 'paradise', 'punishment' etc. associated with emotion-types joy, distress, relief and love. We are assigning a value between -5 to 5 as the concept-valence. We have employed ConceptNet to assign the concept-valence as explained below. Another example that could relate some concepts in a negative manner is *'Disaster'*. *'Disaster'* is an abstract-concept which describes about the properties of massive destructions and loss of lives and it allows key-concepts like, 'cyclone', 'earthquake', 'flood', 'famine', 'tsunami' etc. linked with emotion-types distress, fear, disappointment and hate. At present we have listed 40 such abstract-concepts enlisting about 2000 key-concepts. To process the concept which has not been listed in our database, we employ ConceptNet to find the match with the existing or related key-concepts or assign a 'concept-valence' dynamically for the new concept and store in the database. In order to assign 'concept-valence' to a key-concept SenseNet, being a client, invokes ConceptNet server. In the server *DisplayNode( )* function is employed and it returns all the possible semantically connected entities that ConceptNet has found for the input concept. Server then makes two groups of semantic relations; in the first group all the entries for the relations like 'IsA', 'DefinedAs', 'MadeOf', 'PartOf' are collected and the second group enlists the entries for the relations like, 'CapableOf', 'UsedFor', 'CapableOfReceivingAction'. The first list is again searched for any matching concept in the list. If it fails, from the second list which is actually a list of verbs, the first 5 unique verbs or actions are matched with the scored-verb list and an average score for those verbs is retuned as the concept-valence. For example, for the concept 'ticket' the system failed to find in the existing knowledgebase and hence the following two lists are obtained from the server as explained above;

Possible_concept_list = ['receipt', 'reservation', 'little piece', 'piece of paper', 'piece of cardboard', 'little piece of cardboard', 'paper', 'return ticket']

Possible_action_list = ['get person', 'get person into event', 'represent money', 'allow access', 'represent', 'provide access', "say 'admit'", 'enter', 'speed', 'ride train', 'see',] (the list is truncated for space limitation)

In this case SenseNet first tries with the list, *Possible_concept_list* and it fails to assign a value. So the second list, *Possible_action_list*, is processed and from the second list SenseNet returned the value 3.534 by averaging the scores of the verbs, 'allow'; 'access'; 'get'; 'provide' and 'represent'. Hence the value 3.534 is assigned as the concept-valence for the concept 'ticket' and stored in the database for future use.

Domain-knowledge: SenseNet maintains a database of scored named-entities. For example, the score of an entity is stored in the following tuple-like format: *Named-entity [Role, Concept, Genre, General-Sentiment]*, the field 'Role' indicates any of the value from the list {Company, Concept, Country, Object, Other, Person, Product, Service, Team} and the 'Concept' stores a keyword to represent the concept of the entity. Genre indicates any of the 15 genres taken from news domains. The General-Sentiment field indicates either a negative (-1) or positive (+1) impression towards the named-entity. Two examples are given, *Discovery {Object, Rocket, Space News, 1}*, and *Microsoft {Company, Research, Science, 1}*. The value for General Sentiment is a subject to personal-view or opinion. But in general we assigned negative values for those entities that are usually associated with wars, crime or negative concept.

2.5 Assumptions

The rules and algorithms of SenseNet are based on the following assumptions.

Assumption 1: *A concept or named-entity has a valence.* SenseNet maintains a growing list of concepts scored with the help of ConceptNet. A named-entity can be represented by its' type and valence can be calculated by considering the valence of the role and general sentiment. For example, the sentence *"Nearly a year after Katrina flooded New Orleans, the city still does not have a plan for rebuilding"*, the valence of 'Katrina' is set according to the concept-valence of "Cyclone" (-4.433) and moreover the general sentiment (-1) validates the negative polarity of the assigned valence.

Assumption 2: *An action may be associated with more than one concept.* For example, the input sentence, "He studied both medicine and psychology." We have the action word "study" associated with the two concepts 'medicine' and 'psychology'. Hence we have two senses to be assessed: [study, medicine] and [study, psychology]"

Assumption 3: *An action also has a negative, positive or neutral valence represented by a numerical value.* In the database of verb list each verb has a value ranging from -5 to 5. '*Be'* verbs (e.g. is, am, has etc.) are considered neutral assigning 0 as the valence.

Assumption 4: *The smallest unit of the SenseNet processing element is a sense-unit and the core element is a verb accompanied by a concept.* A valid sense-unit must have a verb and a concept associated with that verb. If a verb has a missing concept, a (dummy) positive concept is imagined to form a sense-unit. So a 'sense' can be formed by a sense-unit with or without a subject and associated adjective/attributes.

Assumption 5: *A sense-unit outputs either a negative, positive or neutral valence.* A sense-unit outputs either a negative, positive or neutral valence. For the input, 'President Bush called the space shuttle Discovery on Tuesday to wish the astronauts well, congratulate them on their space walks and invite them to the White House.' The sense-units are: [call, Discovery], [wish, astronauts], [congratulate, them] and [invite, them]. The rules to assign the polarity sign of sense-unit are:

- Neg. Verb + Pos. Concept→ Neg. Polarity (e.g. quit job, kill civilians)
- Neg. Verb + Neg. Concept→ Pos. Polarity (e.g. quit drugs, kill insects)
- Pos. Verb + Pos. Concept→ Pos. Polarity (e.g. buy car, fly Europe)
- Pos. Verb + Neg. Concept → Neg. Polarity (e.g. buy gun, encourage terrorist)

The valence is calculated by adding the scores of both verb and concept.

Assumption 6: *An adjective (having positive or negative value) can influence the intensity or polarity sign of the sense-valence.* As examples, "I like movies.", and "I like romantic movie"; if the valence of positive sense of the sense-unit ("like-movie") for the first sentence is 7.56, the valence of sense will be intensified for the second one because of positively scored (5.00) concept-modifier 'romantic'. Similarly the valence of the negative sense of the sentence "I dislike this camera." is higher than that of the sentence "I dislike this broken camera" due to the negative score (-2.692) for the adjective 'broken'. The polarity sign of the sense-valence is toggled by the adjectives when there is a negative scored adjective qualifying a positive sense-unit. For example, for these two sentences, "I usually take photos." and "I usually take bad photos."; if the sense-valence of the first one is +7.54, the sense-valence for the second one will be –7.54 for the negative scored (-4.281) adjective 'bad' (with the positive sense-unit 'take-photo'). In the case of negative sense-unit both positive and negative scored adjectives intensify the sense-valence only. Thus the negative valence for the sentence "The attack killed three innocent civilians" is higher than that of the sentence "The attack killed three civilians" for the positively scored (3.571) adjective 'innocent'.

Assumption 7: *An adverb (having positive or negative value) can influence the intensity or polarity sign of the sense-valence.* The negatively score adverb with a negative scored action toggles the polarity sign of the sense-valence. For example, "I miss the morning lectures." is assessed negatively (-8.324) for sense-unit ['miss-lectures'] and for the sentence, "I often miss morning lectures." the valence is intensified by a unit factor (-9.324) for the positive scored (4.334) adverb ‘often’ but the polarity sign of the sense-valance (-8.324) for the sentence “I hardly miss the morning class” is toggled to indicate the positive valence (8.324) of the sentence due to negatively scored (-4.166) adverb ‘hardly’. Similarly the positive score (7.184) for [‘complete’, ‘assignment’] for the sentence “I completed the assignment on time” is set to negative (-7.184) for the sentence “I rarely completed the assignment on time”. This type of sentiment is classified as positively manner negative sentiment.

Assumption 8: *The concept-valence of an actor may modify the polarity-sign of a sense-valence.* Pronouns (e.g. I, he, she etc.) and proper names (not found in the listed named-entity) are considered positively valenced concepts as actors and this value does not make any exception to the rules to assign the polarity-sign of valence but a negatively valenced concept as an actor might be handled with some exceptions. For example, the input sentence “The robber arrived with a car and mugged the store-keeper.” gives a negative valence for the actor (robber) and produces positive and negative assessment for sense-units [“arrive, car”] and [“mug, store-keeper”] respectively. In such cases where a negative-role actor is associated with both positive and negative sense-unit, the positive sense-valence is toggled to negative and thus a totally negative sense-valence is set. But if a negative scored actor is associated with all positively scored sense-unit(s) the polarity sign is not toggled and valance of the positive score is increased by a unit. For following example, “The kidnapper freed the hostage” gives a positive score (6.583) for the sense-unit [‘free, hostage’] connected with negatively scored (-2.334) actor ‘kidnapper’. In this case SenseNet will output 7.583 as the sense-valence which implies that an action done by a negative-role actor is not necessarily always negative.

Assumption 9: *The valence of a ‘Concept’ not in the list can be assigned by considering the valences of the action(s) that are possibly performed by that concept.* It is tedious to enlist all the key-concepts and Abstract-Concept because the list might be too long. If a concept is not found in the database, ConceptNet 2.0 is employed and the detail process is already explained in the section 2.4.

Assumption 10: *Two interrelated negative senses make a positive sense*. If there are two negatively valenced senses dependent with each other (i.e. dependant clauses), that two senses are united and the polarity sign is changed to positive. This type of sentiment is classified as negatively manner positive sentiment. For example, the sentence, “It is difficult to take bad picture with this camera.” outputs two interrelated senses [‘it’, ‘is’, ‘difficult’] and [‘camera’, ‘take’ ‘bad-picture’] and both produces negative sense-valences (-10.00 and -11.945) but the final valence is set to positive (10.972) for the sentence.

Assumption 11: *Negation and Conditionality.* If a sense has a negation with the verb, the polarity-sign is toggled. For the sentence “Bill did not sign the agreement” the sense-valence (10.618) for [‘Bill’, ‘sign’, ‘agreement’] is changed to negative for the negation. The conditionality (e.g. if ... then) is handled by the language parser to set the attribute of the associated verb(s) (i.e. conditional, confirmed or not confirmed etc.). If a sense has a verb with the attribute ‘conditional’, the sense-valence is set neutral (0.0) for that sense. Thus for the sentence “If I win the lottery, I will give a party”, the SenseNet outputs 0 instead of 11.534 (the average sense-valence for the senses [‘I’, ‘win’, ‘lottery’] and [‘I’, ‘give’, ‘party’])

Assumption 12: *The average value of absolute sense-valences of a sentence, S, is the sentence-score of that sentence.* If a sentence, S has N many sense-tuples, the sentence-score of the sentence, S is assigned according to the given formula;

|Sentence-Score(S)| = average (abs (sense*1*_valence)+…abs(sense*N*_valence))..(4)

The polarity sign of the sentence-score is set according to the sign of the sense-valence which value is the maximum among the sense-valences of that sentence. A signed Sentence-Score is termed as Sentiment-Score that varies between -15 to 15. The Average of sentiment-score(s) is considered if there is more than one sentence in the input texts.

2.6 SenseNet Browser

SenseNet browser (Figure 2 shows an example) is the main user interface of the system that allows a user to input a chunk of texts and visualize sentiment of each line of the input text. Clicking on the individual symbol of ‘sentiment-view’ pane will display the valences of sense(s) that corresponds to the symbol and particular sentence of the text-chunk. At present the browser visualizes five kinds of sentiments namely, positive, negative, neutral, positively manner negative and negatively manner positive sentiments represented by different kinds of circles with plus and minus signs. Each circle represents the underlying sentiment of a particular sentence.

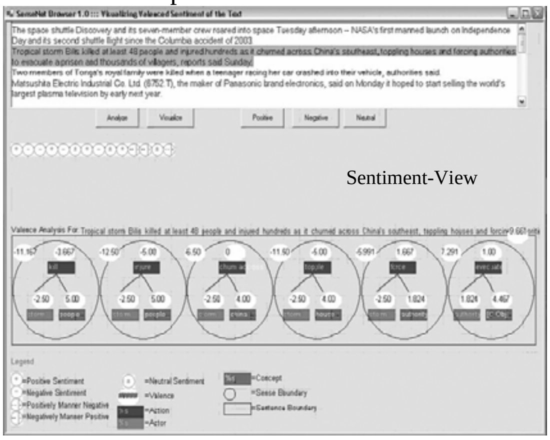

Figure 2 SenseNet browser; the user interface of SenseNet

3 Test and Evaluation

We have tested our system to analyze sentence-level sentiment and matched the score with human-decided score. And we have obtained about 90% accuracy for sensing sentence-level positive and negative sentiment. For details see Survey report.

4 Conclusion

In general terms the research aims at giving computer programs a skill known as *emotional intelligence*, including the ability to recognize, model, and understand human emotion, to convey emotion, and to respond to it appropriately. The linguistic approach to sentiment-sensing from texts would strengthen human-computer interaction with fun. We plan to implement a user interface to set user-specific preferences (e.g. personal opinion about particular entities) that might help the system to perform better based on several user-models.

Reference

Alessandro Valitutti, Carlo Strapparava, and Oliviero Stock. 2004. Developing Affective Lexical Resources, *PsychNology Journal*, 2 (1): 61-83.

Anthony Boucouvalas and Xu Zhe. 2002. Text-to emotion engine for real time internet communication. In *Proceedings of International Symposium on CSNDSP*, pages 164-168, Staffordshire University, UK, July.

Christiane Fellbaum (Ed). 1999. *WordNet: An Electronic Lexical Databases*, MIT Press, Cambridge, Massachusetts.

Clark Elliott. 1994. Components of two-way emotion communication between humans and computers using a board, rudimentary model of affect and personality. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 1(2):16-30.

Connexor Oy. 2005. *Machinese Syntax*, web-site: http://www.connexor.com/connexor/

Fabrizio Sebastiani. 2002. Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1):1-47, March.

Fiorella de Rosis and Floriana Grasso. 2000. Affective text generation, In Paiva, A. (ed): *Affective Interactions*, Lecture Notes in Artificial Intelligence, 1814: 204-218, Springer-Verlag.

Hugo Liu and Push Singh. 2004. ConceptNet: A Practical Commonsense Reasoning Toolkit, *BT Technology Journal*, 22(4):211-226, Kluwer Academic Publishers.

Hugo Liu, Henry Lieberman, and Ted Selker. 2003. A Model of Textual Affect Sensing using Real-World Knowledge, In *Proceedings of the Seventh International Conference on Intelligent User Interfaces (IUI 03)*, pages 125-132, Miami, USA, ACM.

James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 2001. Linguistic inquiry and word count: *LIWC* (2nd ed.) [Computer software]. Mahwah, NJ: Erlbaum.

Janyce Wiebe, Rebecca Bruce, Matthew Bell, Melanie Martin, and Theresa Wilson. 2001. A Corpus Study of Evaluative and Speculative Language. In *Proceedings of the 2nd ACL SIGdial Workshop on Discourse and Dialogue*, pages 221-230, Aalborg, Denmark, September.

Nuno Maria and Mário J. Silva. 2001. Theme-based Retrieval of Web News. In *Proceedings of the third International Workshop on the Web and Databases*, Athens, Greece, Published as Springer Lecture Notes in Computer Science, 1997: 26-37.

Survey report, http://ita.co.jp/research/survey/

Tetsuya Nasukawa and Jeonghee Yi. 2003. Sentiment Analysis: Capturing Favorability Using Natural Language Processing, In *Proceedings of the international conference on Knowledge capture*, pages 70-77, ACM Press.