Emotion Holder for Emotional Verbs – The role of Subject and Syntax

Dipankar Das and Sivaji Bandyopadhyay

Department of Computer Science and Engineering,
Jadavpur University, Kolkata, India
dipankar.dipnil2005@gmail.com, sivaji_cse_ju@yahoo.com

Abstract. Human-like holder plays an important role in identifying actual emotion expressed in text. This paper presents a baseline followed by syntactic approach for capturing emotion holders in the emotional sentences. The emotional verbs collected from WordNet Affect List (WAL) have been used in extracting the holder annotated emotional sentences from VerbNet. The baseline model is developed based on the *subject* information of the dependency-parsed emotional sentences. The unsupervised syntax based model is based on the relationship of the emotional verbs with their argument structure extracted from the *head* information of the chunks in the parsed sentences. Comparing the system extracted argument structure with available VerbNet frames' syntax for 942 emotional verbs, it has been observed that the model based on syntax outperforms the baseline model. The precision, recall and F-Score values for the baseline model are 63.21%, 66.54% and 64.83% and for the syntax based model are 68.11%, 65.89% and 66.98% respectively on a collection of 4,112 emotional sentences.

Keywords: Emotion Holder, VerbNet, Emotional Verb, Subject, Syntax, Arguments.

1 Introduction

In psychology and common use, emotion is an aspect of a person's mental state of being, normally based in or tied to the person's internal (physical) and external (social) sensory feeling [1]. The determination of emotion expressed in the text with respect to reader or writer is itself a challenging issue. Emotion holder extraction research is important for discriminating between emotions that are viewed from different perspectives [2]. A wide range of Natural Language Processing (NLP) tasks such as tracking users' emotion about products or events or about politics as expressed in online forums or news, to customer relationship management are using emotional information. So, the determination of emotion holder from the text invokes a challenge and helps us track and distinguish user's emotion separately.

In linguistics, a grammatical agent or holder is the participant of a situation that carries out the action and also, *agent* or *holder* is the name of the *thematic role*. The basic clue for identifying the emotion holder is the presence of any emotional verb or the appearance of non-emotional verb with any emotional phrases. The argument-based

A. Gelbukh (Ed.): CICLing 2010, LNCS 6008, pp. 385–393, 2010.

relationship of the emotional verbs with other component phrases in an emotional sentence gives the information to tag emotion holder both syntactically and semantically.

The present work aims to identify the emotion holder using two different approaches. A baseline system is developed based on the *subject* information of the emotional sentences parsed using Stanford Dependency Parser [3]. The *precision*, *recall* and *F-Score* values of the holder identification system are 63.21%, 66.54% and 64.83% respectively for the baseline approach.

Another way to identify emotion holder is based on the syntactical argument structure of the emotional sentences corresponding to the emotional verbs. Emotional verbs corresponding to Ekman's six different emotion types are retrieved from the WordNet Affect Lists (WAL) [15]. A total of 4,112 emotional sentences for these 942 emotional verbs have been extracted from the English VerbNet [4]. The holder related information as specified in the VerbNet such as *Experiencer, Agent, Actor, Beneficiary* etc. are properly tagged in the correct position of the syntactical frames for each sentence. All possible subcategorization frames and their corresponding syntaxes, available in the VerbNet are retrieved for each emotional verb. To achieve the objective, the *head* of each chunk is extracted from the dependency-parsed output. This chunk level information helps in constructing the syntactic argument structure with respect to the key emotional verb. The acquired syntactic argument structure is mapped to all the possible syntax structures present for each emotional verb in the VerbNet. If the syntactic argument structure of a sentence matches with any of the syntax structures extracted from the VerbNet for each emotional verb, the holder role associated with the VerbNet syntax is then assigned the holder tag in the appropriate component position of the syntactical arguments. Two separate techniques have been adopted for extracting the argument structure. One is from parsed result directly and another is from the corpus that has been POS tagged and chunked separately. The *precision* (P) and *recall* (R) values of these two techniques are 68.11 % (P), 63.04% (P) and 65.89 % (R), 64.34% (R) respectively on a collection of 4,112 emotional sentences. But, it has to be mentioned that the first technique gives significantly better *F-Score* value (66.98%) than the second one (62.39%) as the second one fails to disambiguate mostly the arguments from adjuncts. So, the dependency parser based method has been selected for emotion holder identification task. It has been observed that the baseline model suffers from the inability to identify emotion holder from the sentences containing passive senses. Although the *recall* value has been decreased in the syntactic model, it outperforms over the baseline model significantly in terms of *precision* and *F-Score*.

The rest of the paper is organized as follows. Section 2 describes the related works done in this area. The baseline system based on parsed data is described in Section 3. Two methods for developing syntax based model for holder identification is discussed in Section 4. Evaluation mechanism along with associated results is specified in Section 5. Finally Section 6 concludes the paper.

2 Related Work

Identification of opinion with its holder and topic from online media text using semantic role labeling is described in [6]. Importance of verb classes and linguistic features in classifying *polarity* and *subjectivity* are explained in [20]. Other related works are [7, 8] where they use the named entities to identify opinion holders.

An anaphor resolution based opinion holder identification method exploiting lexical and syntactic information from online news documents is narrated in [9]. Using generated features for each named entity and sentence pair, the machine learning based classification task for "not holder", "weak holder", "medium holder", or "strong holder" from the MPQA corpus is carried out in [10]. Identifying opinion holders for Question Answering in opinion text and the supporting annotation task are reported in [12].

The work on labeling the arguments of the verbs with their semantic roles using a novel frame matching technique is described in [21]. The present work is mostly related to the work described in [21]. But, irrespective of assignment of semantic roles, a technique has been designed to acquire argument structure of a sentence corresponding to the emotional verbs and map them on the frame syntax available in VerbNet for those verbs.

Based on the traditional perspectives, a new emotion holder model [11] is generated containing an emotion knowledge base for emotion expression followed by performing emotion reasoning algorithm and finally implementing the emotions treatment. The identification of the opinion propositions and their holders mostly for verbs is described in [13]. This work is similar to the present approach. But, the application of argument structure to identify emotion holder with respect to emotional verb in this present task signifies the difference from this approach. The comparative study of *subject* based holder identification task with syntax-based technique adopted in this present task is contributory to the platform of emotion holder identification.

3 Subject Based Baseline Model

The emotion holder present for an emotional verb in a sentence is crucial from the perspective of active and passive forms of the sentence. Before going into the detail exploration of the systems, the preparation of the holder annotated gold standard emotional corpus is first described. This is followed by the baseline methodology to extract holder information from parsed sentences.

3.1 Corpus Preparation

The sentiment lexicon, *SentiWordNet* [14] and emotion word lists like *WordNet Affect lists* (WALs) [15] are available in English. The English WALs, based on Ekman's [16] six emotion types are updated with the synsets retrieved from the English *SentiWordNet* to make adequate number of emotion word entries [17]. The list of verbs that have been collected from these six modified WALs, are termed as emotional verbs.

The enlisted emotional verbs are searched through the VerbNet classes. As the member verbs in any VerbNet class share the same syntactic and semantic information, the sentences described in that class is common for all members in that class. The sentences present in a VerbNet class and shared by different members are similar for an emotional verb if it is a member of that class and it can be considered that these sentences carry emotion due to the presence of such emotional verb(s).

If an emotional verb is found in any of the member verbs of any VerbNet class, the sentences corresponding to that class have been retrieved to construct the emotion corpus. The holder related tags (e.g. *Agent, Experiencer, Beneficiary* and *Actor*) are used to tag the retrieved sentences accordingly to prepare the gold standard holder annotated corpus. Table 1 shows the detailed statistics of the Emotion Corpus. Out of total 5,432 retrieved emotional sentences, 4,112 sentences are tagged with their holder related tags accordingly.

Table 1. Statistics of the Emotion Corpus

Information	# Total items found
Emotional Verbs in WordNet Affect list	1,278
Emotional verbs present in VerbNet	942
Retrieved Emotional Sentences	5,432
Annotated emotional sentences with *Holder* tag	3,156
Annotated sentences with other tag (*Experiencer/ Beneficiary* etc.)	956
Distinct VerbNet classes	523

3.2 Dependency Parsing and Subject Extraction

Stanford Parser [3], a probabilistic lexicalized parser containing 45 different part of speech (POS) tags of Pen Tree bank has been used to get the parsed sentences with dependency relations. The input emotional sentences are passed through the parser. The dependency relationships extracted from the parsed data are checked for predicates "*nsubj*" so that the *subject* related information in the "*nsubj*" predicate is considered as the probable candidate for identifying the emotion holder. Other dependency relations are filtered out from the parsed output. The present baseline system is developed based on the filtered subject information only. An example sentence is noted below whose parsed output and dependency relations are shown in Table 2. Here, the "*nsubj*" relations containing the emotional verb "grieve" tags "I" as an emotional holder.

*"I **grieve** for my departed Juliet."*

Table 2. Parsed Results

Parsed Output	*Dependency Relation*

(ROOT
  (S
    (**NP** (PRP I))
    (**VP** (VBP grieve)
      (**PP** (IN **for**)
        (**NP** (PRP$ my)
(JJ departed)(NN Juliet))))
    (. .)))
	nsubj(**grieve**-2, **I**-1)
	poss(Juliet-6, my-4)
	amod(Juliet-6, departed-5)
	prep_for(grieve-2, Juliet-6)

This baseline model is evaluated on the gold standard holder annotated emotional sentences that has been extracted from VerbNet. Total 4,112 sentences are evaluated and evaluation results are presented in Table 3 in terms of *precision*, *recall* and *F-Score*. It has been observed that the grammatical holders of the emotional sentences containing passive sense are often confused with the *subject* information. So, the next step is to explore the syntactical way for identifying argument structure of the sentences for their corresponding emotional verbs and to capture the emotion holder as a *thematic role* respectively.

4 Syntax Based Model

The syntax of a sentence is an important clue to capture the holder inscribed in text. More specifically, the argument structure or subcategorization information for a verb plays an essential role to identify the emotion holder from an emotional sentence. A subcategorization frame is a statement of what types of syntactic arguments a verb (or an adjective) takes, such as objects, infinitives, that-clauses, participial clauses, and subcategorized prepositional phrases [18]. VerbNet (VN) [4] is the largest online verb lexicon with explicitly stated syntactic and semantic information based on Levin's verb classification [23]. It is a hierarchical domain-independent, broad-coverage verb lexicon with mappings to other lexical resources such as WordNet [24], XTAG [25] and FrameNet [22]. Irrespective of other well-known lexical resources, VerbNet is used throughout this experiment as the main thrust for identifying the emotion holders is based on the characteristics of the emotional verbs only.

The existing syntax for each emotional verb is extracted from VerbNet and a separate rule based argument structure acquisition system is developed in the present task for identifying the emotion holder. The acquired argument structures are compared against the extracted VerbNet frame syntaxes. If the acquired argument structure matches with any of the extracted frame syntaxes, the emotion holder corresponding to each emotional verb is tagged with the holder information in the appropriate slot in the sentence.

4.1 Syntax Acquisition from VerbNet

VerbNet associates the semantics of a verb with its syntactic frames and combines traditional lexical semantic information such as thematic roles and semantic predicates, with syntactic frames and selectional restrictions. Verb entries in the same VerbNet class share common syntactic frames, and thus they are believed to have the same syntactic behavior. The VerbNet files containing the verbs with their possible subcategorization frames and membership information are stored in XML file format. E.g. the emotional verbs "*love*" and "*enjoy*" are member of the *admire-31.2-1* class and "*enjoy*" also belongs to the class *want-32.1-1*. A snapshot of the XML file for the *admire-31.2-1* class is given below.

…<VNCLASSID="admire-31.2"
… <SUBCLASSES>….
<VNSUBCLASS ID="admire-31.2-1">
<MEMBERS>….
<MEMBER name="**love**" wn="love%2:37:00 love%2:37:02 love%2:37:01"/>
<MEMBER name="**enjoy**" wn="enjoy%2:37:00 enjoy%2:37:01 enjoy%2:34:00"/>…..
<THEMROLES/> <FRAMES>
<FRAME> <DESCRIPTION descriptionNumber="8.1" primary="TO-INF-SC" secondary="" xtag="0.1"/> …. <EXAMPLE>I loved to write.</EXAMPLE>
<SYNTAX> <NP value="Experiencer"> <SYNRESTRS/> </NP> <VERB/> <NP value="Theme">
<SEMANTICS> <PRED value="**emotional_state**">
<ARGS> <ARG type="Event" value="E"/> <ARG type="**VerbSpecific**" value="**Emotion**"/> <ARG type="ThemRole" value="Experiencer"/> …..
</ARGS> </PRED> </SEMANTICS> </FRAME>…..

The XML files of VerbNet are preprocessed to build up a general list that contains all member verbs and their available syntax information retrieved from VerbNet. This preprocessed list is searched to acquire the syntactical frames for each emotional verb. One of the main criteria considered for selecting the frames is the presence of "*emotional_state*" type predicate associated with the frame semantics.

4.2 Argument Structure Acquisition Framework

To acquire the argument structure for a sentence, two separate approaches, Methods A and B, have been used, one (Method A) is from the parsed result directly and another (Method B) is from the POS tagged and chunked sentences accordingly.

The parsed emotional sentences are passed through a rule based *phrasal-head* extraction process to identify the phrase level argument structure of the sentences corresponding to the emotional verbs. The extracted *head part* of every phrase from the well-structured bracketed parsed data is considered as the component of the argument structure. For example, the *head* parts of the phrases are extracted to make the phrase level pattern or argument structures of the following sentences.

Sentence1: "*Caesar killed Brutus with a knife.*"
Parsed Output:

(ROOT (S (**NP** (NNP Caesar)) (**VP** (VBD killed) (**NP** (NNS Brutus)) (**PP** (IN **with**) (NP (DT a) (NN knife)))) (. .)))

Acquired Argument Structure: [NP VP NP PP-with]
Simplified Extracted VerbNet Frame Syntax: [<NP value="*Holder*"> <VERB/> <NP-*patient*> <PREP value="*with*">]

Sentence2: "***I love*** *everybody*."
Parsed Output:

```
(ROOT (S (NP (PRP I))(VP (VBP love))(NP (NN everybody))) (. .))
```

Acquired Argument Structure: [NP VP NP]
Simplified Extracted VerbNet Frame Syntax: [<NP value="*Experiencer*" ></VERB><NP-*theme*>]

Sentence3: "*The children liked that the clown had a red nose.*"
Parsed Output:

(ROOT (S (**NP** (DT The) (NNS children)) (**VP** (VBD liked) (**SBAR** (IN **that**)
(S (NP (DT the) (NN clown)) (VP (VBD had) (NP (DT a) (JJ red) (NN nose)))))) (. .)))

Acquired Argument Structure: [NP VP SBAR-that]
Simplified Extracted VerbNet Frame Syntax: [<NP value="*Experiencer*"><VERB/><NP-theme><SYNRESTR type="*that*_comp"/>

It is to be mentioned that, the phrases headed by "S" (sentential complement), "PP" (Preposition Phrase), "NP" (Noun Phrase) and followed by the emotional verb phrase contribute in structuring the syntactical argument. One tag conversion routine has been developed to transform the POS information of the system-generated argument structure for comparison with the POS categories of the VerbNet syntax. It has been observed that the phrases that start with ADJP, ADVP (adjective, adverbial phrases) tags generally do not contribute towards valid argument selection strategy. But, the entities in the slots of active frame elements are added if they construct a frame that matches with any of the extracted frames from VerbNet. The *head* part of each phrase with its component attributes (e.g. "*with*" component attribute for "PP" phrase) in the parsed result helps in identifying the maximum matching possibilities.

Another alternative way to identify the argument structure from a sentence is carried out based on the POS tagged and chunked data. The emotional sentences are tagged with an open source Stanford Maximum Entropy based POS tagger [5]. The best reported accuracy for the POS tagger on the Penn Treebank is 96.86% overall and 86.91% on previously unseen words. The POS tagged sentences are passed through a Conditional Random Field (CRF) based chunker [19] to acquire chunked data where each component of the chunk is marked with *beginning* or *intermediate* or *end* corresponding to the elements slot in that chunk. The POS of the *beginning* part of every chunk has been extracted and frames have been developed to construct the argument structure of the sentence corresponding to the emotional verb. The acquired argument structure of a sentence is mapped to all of the extracted VerbNet frames. If a single match is found, the slot devoted for the holder in VerbNet frame is used to tag in the appropriate slot in the acquired frame. For example, the argument structure acquired from the following chunked sentence is "NP-VP-NP".

```
I/PRP/B-NP love/VBP/B-VP them/PRP/B-NP ././O
```

But, it has been observed that this second system suffers from the inability to recognize arguments from adjuncts as the system blindly captures *beginning* parts as arguments whereas they are adjuncts in real. So, this system is biased to the *beginning* chunk.

5 Evaluation

The evaluation of the baseline system is straightforward. The emotion holder annotated sentences are extracted from the VerbNet and the sentences are passed through the baseline system to annotate the sentences with their *subject* based holder tag accordingly. A total of 4,112 sentences are evaluated and the *precision*, *recall* and *F-Score* values are shown in Table 3. It is observed that the *subject* information helps in identifying emotion holder with high *recall*. But, the holder identification task for passive sentences fails in this baseline method and hence there is a fall in *precision* value. Two types of unsupervised rule based methods have been adopted to acquire the argument structure from the emotional sentences. It has been observed that, the Method-A that acquires argument structure from parsed result directly outperforms the Method-B that acquires these structures from POS tagged and chunked data. The *recall* value has decreased in Method-B as it fails to distinguish the arguments from the adjuncts. The emotion holder identification system based on argument structure directly from parsed output gives satisfactory performance.

Table 3. Precision, Recall and F-Score values of the Baseline and Syntactic model

Type	*Baseline Model* (in %)	*Syntactic Model* (in %)	
		Method-A	Method-B
Precision	63.21	68.11	63.05
Recall	66.54	65.89	64.34
F-Score	64.83	66.98	62.39

6 Conclusion

In this work, the emotion holder identification task is carried out based on the roles associated to *subject* information. The syntactic way of developing the holder extraction module by focusing on the role of arguments of the emotional verbs improves the result significantly. Further works need to be on sentences with non-emotional verbs but emotional phrases. The holder-annotated corpus preparation from VerbNet especially for emotional verbs followed by the argument extraction module can be further explored through the help of machine learning approach.

References

1. Zhang, Y., Li, Z., Ren, F., Kuroiwa, S.: A Preliminary Research of Chinese Emotion Classification Model. IJCSNS International Journal of Computer Science and Network Security 8(11), 127–132 (2008)
2. Seki, Y.: Opinion Holder Extraction from Author and Authority Viewpoints. In: SIGIR 2007. ACM, New York (2007), 978-1-59593-597-7/07/0007
3. de Marneffe, M.-C., MacCartney, B., Manning, C.D.: Generating Typed Dependency Parses from Phrase Structure Parses. In: 5th International Conference on Language Resources and Evaluation (2006)

4. Kipper-Schuler, K.: VerbNet: A broad-coverage, comprehensive verb lexicon. Ph.D. thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia, PA (2005)
5. Manning, C.D., Toutanova, K.: Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. In: Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, EMNLP/VLC (2000)
6. Kim, S.-M., Hovy, E.: Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. ACL (2006)
7. Kim, S.-M., Hovy, E.: Identifying Opinion Holders for Question Answering in Opinion Texts. In: Proceedings of AAAI 2005 Workshop on Question Answering in Restricted Domains (2005)
8. Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. In: Proceedings of HLT/EMNLP 2005 (2005)
9. Kim, Y., Jung, Y., Myaeng, S.-H.: Identifying Opinion Holders in Opinion Text from Online Newspapers. In: 2007 IEEE International Conference on Granular Computing, pp. 699–702 (2007), doi:10.1109/GrC.2007.45
10. Evans, D.K.: A low-resources approach to Opinion Analysis: Machine Learning and Simple Approaches. NTCIR (2007)
11. Hu, J., Guan, C., Wang, M., Lin, F.: Model of Emotional Holder. In: Shi, Z.-Z., Sadananda, R. (eds.) PRIMA 2006. LNCS (LNAI), vol. 4088, pp. 534–539. Springer, Heidelberg (2006)
12. Wiebe, J., Wilson, T., Cardie, C.: Annotating expressions of opinions and emotions in language. Language Resources and Evaluation 1(2) (2005)
13. Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., Jurafsky, D.: Automatic Extraction of Opinion Propositions and their Holders. In: AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications (2004)
14. Esuli, A., Sebastiani, F.: SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In: LREC 2006 (2006)
15. Strapparava, C., Valitutti, A.: WordNet-Affect: an affective extension of WordNet. In: Proceedings of the 4th International Conference on Language Resources and Evaluation, pp. 1083–1086 (2004)
16. Ekman, P.: Facial expression and emotion. American Psychologist 48(4), 384–392 (1993)
17. Das, D., Bandyopadhyay, S.: Sentence Level Emotion Tagging. In: ACII 2009. IEEE, Los Alamitos (2009)
18. Manning, C.D.: Automatic Acquisition of a Large Subcategorization Dictionary from Corpora. In: 31st Meeting of the ACL, Columbus, Ohio, pp. 235–242 (1993)
19. Phan, X.-H.: CRFChunker: CRF English Phrase Chunker. In: PACLIC 2006 (2006)
20. Vincent, B., Xu, L., Chesley, P., Srhari, R.K.: Using verbs and adjectives to automatically classify blog sentiment. In: Proceedings of AAAI-CAAW 2006, the Spring Symposia (2006)
21. Swier, R.S., Stevenson, S.: Unsupervised Semantic Role Labelling. In: Proceedings of EMNLP (2004)
22. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The Berkeley FrameNet project. In: COLING/ACL 1998, Montreal, pp. 86–90 (1998)
23. Levin, B.: English Verb Classes and Alternation: A Preliminary Investigation. The University of Chicago Press, Chicago (1993)
24. Miller, G.A.: WordNet: An on-line lexical database. International Journal of Lexicography 3(4), 235–312 (1990)
25. XTAG Research Group.: A lexicalized tree adjoining grammar for English. IRCS. University of Pennsylvania (2001)