On the Effectiveness of Emotion Extraction Techniques

Yashar Moshfeghi
School of Computing Science
University of Glasgow
Glasgow, UK
Yashar.Moshfeghi@glasgow.ac.uk

Joemon M. Jose
School of Computing Science
University of Glasgow
Glasgow, UK
Joemon.Jose@glasgow.ac.uk

ABSTRACT

Emotion is considered to be an important factor in human decision making and consciousness. Due to its perceived importance, recently a number of works try to explore emotion in IR tasks. However, the complexity of emotion extractors and lack of understanding on their effectiveness hinder the progress of such works. In this paper, we conduct a comparative study on the effectiveness of three emotion extractors. Our findings show a superior performance of an extractor based on OCC model.

Categories and Subject Descriptors: H.3.3 Information Storage and Retrieval - *Information Search and Retrieval - Information Filtering*; H.3.4 Information Storage and Retrieval - *Systems and Software - Performance Evaluation*

General Terms: Performance, Experimentation

Keywords: Emotion Extraction System, Text, Evaluation, OCC Model

1. INTRODUCTION

With increased usage of the internet in everyday life, there has been a proliferation of data in varying format from many sources, including news and user-generated contents such as reviews, social networks, twitter, etc. As a consequence, there is a need for exploring new features such as sentiment and emotion which may help to better differentiate between the data thereby improving the information retrieval effectiveness [1]. Sentiment analysis is used for opinion mining applications and the purpose is to use the popularity of opinions to support information retrieval, seeking and mining tasks. In a similar way, there has also been an increase in research activity on emotion extraction and its use in IR applications [2]. Although emotion is subjective, it is presented in some objectively deducible ways in written documents. A careful choice of words can deliver a particular emotion from one person to another. The fact that natural language can convey emotion fascinated linguists and psychologists [3]. In recent years there has been an increasing amount of research in both academia and practice to enable computers to extract emotion from textual resources [3]. This is due to the importance and usefulness of the applications utilising these techniques in the areas such as affective computing, opinion mining, market analysis and human computer interaction.

There are multiple views of on what emotion is and how it is to be represented [4]. Ekman [5] regards emotion as psychosomatic states and categorises them into six discreet categories[1], and some commercial systems follow this approach [6]. On the other hand, the OCC model[2] categorises emotion into 24 states [7]. This approach is considered to be superior by the cognitive psychological community, and is relied upon by the state-of-the-art research on emotion extraction by Masum et al. [3]. An argument is made for incorporating emotion in IR tasks and some initial work has been done in this field [2], overall the developments are in a nascent state. Hindering progress is the complexity of emotion extraction techniques and lack of understanding on their effectiveness.

A sophisticated emotion extractor system is the work presented by Liu et al. [8][3]. This approach uses a textual affective sensing engine which utilise common-sense knowledge to classify texts into six basic Ekman emotions. Common-sense knowledge is a graph where real-life concepts are the nodes and their relationships are the edges of the graph. The problem of this approach is that the linguistic facet of the sentence is not taken into account. Therefore, sentence like "it is impossible to cook a bad meal following this recipe" and "I will cook a bad meal following this recipe" will be categorised in the same group. In addition, it has been argued that using the six categories to classify text is not optimal, as they are expression-based emotion and do not consider the cognitive aspect of emotion such as belief, decision and intention [3].

Synesketch[4] uses natural language processing techniques

OAIR'13, May 22-24, 2013, Lisbon, Portugal.

[1]Specifically happiness, sadness, fear, anger, disgust and surprise [5]

[2]The OCC emotion model specifies 22 emotion types and two cognitive states. The OCC emotion categorises as joy , distress, happy-for, sorry-for, resentment, gloating, hope, fear, satisfaction, fears-confirmed, relief, disappointment, shock, surprise, pride, shame, admiration, reproach, gratification, remorse, gratitude and anger. The two cognitive states are love and hate [7].

[3]Liu et al. [8] provide an open source version of their approach called EmpathyBuddy. EmpathyBuddy considered to be the best performing open source emotion extraction system [3].

[4]Synesketch is an open-source Java API for textual emotion recognition. The API is bundled with a number of imaging tools that allows users to automatically create visualisations

to analyse a text, and uses the Affective WordNet lexicon [9] to calculate the emotion weights of words in text. Synesketch categorises emotion from text into six basic Ekman emotions. Masum et al. [3] uses contextual valence values towards triplet parts (i.e. subject, verb, and object) and extract emotion from text using OCC model rules. However, their approach depends on a non-open source triplet extractor. The triplet extractor is a key component for both sentiment analysis and emotion extraction parts. This make it impossible to distribute the software widely. We have re-implemented Masum et al. work, with the generous support from the authors. Finally, Masum et al. conducted only a user based evaluation of their emotion extractor and hence our results could not be compared [3]. This is the first attempt to systematically study the effectiveness of an emotion extractor based on the OCC model.

There is no standard resource nor methodologies for evaluating the usefulness and effectiveness of emotion extractor systems. The evaluation of such systems is performed using either different resources or different metrics and comparisons are impossible to make. For IR tasks, it is important to study the effectiveness of emotion extraction techniques. The objectives of this study is to conduct a comparative study of the approaches so that the effectiveness can be compared. In particular, in addition to Masum et al. [3] system, two publicly available commercial systems, Synesketch [10] and EmpathyBuddy [8] were used. The rest of the paper is organised as follows. Our implementation of Masum et al. approach called "Open-OCC (O-OCC)" and the experimental procedure are explained in Section 2 and 3 respectively. The results are presented and discussed in Section 4, and the paper is concluded in Section 5.

2. IMPLEMENTATION OF O-OCC

Figure 1 shows the architecture of the components involved in our implementation of the Masum et al. [3] approach, namely the prior valence provider, triplet extractor, sentiment analyser, and emotion extractor. The rest of this section explains each component in detail to show the complexity of a state-of-the-art emotion extractor. When appropriated, the differences with the original approach are discussed.

The Prior Valence Provider: This component is responsible for creating and expanding a set of base lists, each of which maps set of words (categorised morphologically, e.g., verbs, adverbs, adjectives, and nouns) to their prior valence (i.e., positive or negative). The initial base lists of verbs, adjectives and adverbs are created with the help of WordNet [11]. A base list for nouns is created with the help of ConceptNet and an initial base list for named entities is also formed using Opinmind (`www.Opinmind.com`). If a prior valence for a term is not available in a base list, the prior valence provider automatically assigns a valence for that word by first obtaining the synonyms of that word using a thesaurus (`www.thesaurus.com`), then screening the synonyms with respect to the corresponding base lists for which numerical values are already assigned, and finally averaging the obtained valence as the valence value of the word. The new word and its valence are then inserted into the base list.

based on textual emotions. More information can be found at www.synesketch.krcadinac.com.

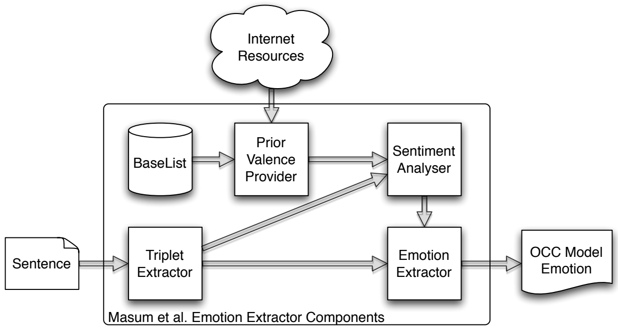

Figure 1: Masum et al. Emotion Extractor Architecture

The Triplet Extractor: A triplet refers to the three-component structure of a sentence: subject, verb and object. Each component may have several attributes called modifier. Examples of modifiers are adjectives, adverbs, and noun phrases. The triplet extractor developed in Masum et al. work depends on a closed-source syntactic parser system. To extract triplets, first a set of triplets for a given sentence is obtained from the Machinese Syntax (`www.connexor.com/connexor/`) (i.e. syntactic parser). Since it is not available, we have to create our own. Hence, a triplet extractor systems was developed, namely Stanford-based triplet extractor. The Stanford-based triplet extractor is based on Stanford parser which generates a highly accurate (86.32%)

```
(ROOT
  (S
    (NP
      (NP (JJ relentless) (NN optimization))
      (PP (IN of)
        (NP (NN information) (NN retrieval) (NN effectiveness))))
    (VP (VBZ has)
      (VP (VBN driven)
        (NP (NN web) (NN search) (NNS engines))
        (PP (TO to)
          (NP
            (NP (JJ new) (NN quality) (NNS levels))
            (SBAR
              (WHADVP (WRB where))
              (S
                (NP (JJS most) (NNS people))
                (VP (VBP are)
                  (VP (VBN satisfied)
                    (NP
                      (NP (JJS most))
                      (PP (IN of)
                        (NP (DT the) (NN time))))))))))))
    (. .)))
```

Deepest Verb

Deepest Sentence

Figure 2: A sample output of the Stanford parser for the sentence: *"relentless optimisation of information retrieval effectiveness has driven web search engines to new quality levels where most people are satisfied most of the time."*

The Stanford-based Triplet Extractor utilises the output of the Stanford parser. Given a sentence, Stanford parser returns a parse tree. Figure 2 shows the parse tree for a sample sentence. To extract triplets from the parse tree, the deepest sentence (DS) in the tree is first found. DS is the sentence which does not have any other sentences under itself in the parse tree (shown in Figure 2). Each DS usually consisting of two main parts, the noun phrase (NP), and verb phrase (VP). The deepest noun in the NP is considered as

[5]Parse tree is a tree that represents the syntactic structure of a sentence based on a formal grammar.

the subject of the DS and other members in the NP such as adjectives and nouns are considered as the attributes of the subject.

The VP usually consist of two parts, namely verb and object. The deepest verb in the VP is considered as the verb of DS, and adverbs in DS are considered as the attributes of the verb. The deepest noun in the VP is considered as the object of the DS. As before, other members in the NP such as adjectives and nouns are considered as the attributes of the object. The information extracted from each DS constitutes a triplet. After that the DS is removed from the sentence and the procedure is repeated until there is no DS left in the sentence. Finally, the dependencies among triplets are taken into account.

The Sentiment Analyser: All the triplets obtained from the input sentences are processed to assign a valence value to the sentence. The valence of the attributes is calculated by averaging the prior valence value of the terms in the attribute set. Then the valence of each part (i.e. subject, verb, or object part) is calculated by applying a set of rules considering the valence value of the actual part (i.e. subject, verb, or object) and its attribute part. Then the valence of a triplet is calculated by first combining the valence of verb and object part (named as verb-object valence) and then combining the valence of subject part with the verb-object part. The combination is based on a set of rules [13]. Finally, the overall valence of the sentence is calculated [13] based on the type of dependencies among triplets.

The OCC Emotion Extractor: The cognitive structure of the OCC model can be characterised by specific rules and their interplay with several variables. There are two kinds of variables involved: emotion inducing variables and emotion intensity variables. Multiple emotions can be inferred from a given situation depending on whether states expressed by certain cognitive variables hold or not hold.

3. EXPERIMENTATION METHODOLOGY

News Headlines Data Set: For our experiments we used the collection *SemEval-2007 Task 14: Affective Text* [14] consisting of a training set which contains 250 headlines and a test set which contains 1000 headlines. Each news headline is annotated with six Ekman emotions. For each emotion, an interval between 0 to 100, where 0 indicates that the emotion is not present in the given headline, and 100 indicates maximum emotional load, is used.

Since there are no publicly available data sets with OCC model emotion categories, we used the above data sets for our experiment. These data sets are suitable since (1) news headlines are intentionally written with an emotionally rich content to provoke readers' attention [14]; (2) the emotion extractor systems that we are examining are sentence-based level; and finally, (3) the outcome of two benchmarks (i.e., Synesketch and EmpathyBuddy) are Ekman emotion categories. The outcome of Masum et al. is not compatible with the Ekman emotion model and there was no guidance in the literature available on how the OCC model's emotions map to Ekman emotions. We therefore came up with a semantic mapping presented in Table 1.

Metrics: Effectiveness of systems were compared using precision, recall, and F-measure. These measures are commonly used in sentiment analysis and emotion extraction literature. For sentiment analysis, the results are calculated

OCC Model Emotion	Ekman Emotion
Surprise, Shock	Surprise
Hate, Anger, Resentment	Anger
Fear, Fears-confirmed	Fear
Sorry-for, Distress, Remorse, Shame	Sadness
Joy, Happy-for, Gloating, Relief, Pride, Admiration, Love, Gratification, Satisfaction, Gratitude, Hope	Joy
Reproach, Disappointment	Disgust

Table 1: Mapping of OCC model emotion to Ekman emotion

for positive sentences, negative sentences and overall. It was important to find an approach that has both the highest overall effectiveness and the best balanced effectiveness in both positive and negative sentences. For emotion extraction, the results are calculated for each individual emotion as well as overall.

Tuning O-OCC: O-OCC is rule-based, and its output is a binary value, corresponding to the dominant emotions in the sentence. In Strapparava and Mihalcea [14], a threshold of 50 is introduced in order to transform the emotion value provided by the judges for each emotion in each news headline sentence to a binary scale. All the emotions with the value in the range of (50, 100] will be considered as 1 and those in the range [0, 50] as 0. However, in our opinion, this threshold is too high, having a negative impact on the performance of O-OCC, and gives us a wrong estimation of the effectiveness of this system. In order to overcome the threshold setting problem, an exhaustive experiment considering all possible thresholds from 0 to 100 was performed. The system that outperforms other approaches for a higher number of thresholds will be considered as the best performing system. Finally, all systems were tuned using the training set and tested on the test set.

Tuning EmpathyBuddy and Synesketch: The output of EmpathyBuddy and Sysesketch is a value from 0 to 1 for each of the Ekman emotions. Since we introduce a threshold on the test set, we also have to optimise the effectiveness of these systems with respect to the threshold. Not defining a cut-off value for each emotion for these systems introduce a bias in the comparison of the systems as the precision value of these systems will be lower than their optimal value. Therefore, it is important to calculate the cut-off value. This should be the value where the system has the highest F-measure. We, thus, selected, for each emotion, a different cut-off value corresponding to the highest F-measure value on the training set.

4. RESULTS

An evaluative comparison between the Synesketch, EmpathyBuddy and O-OCC emotion extractors across all emotions is presented in Figure 3, 5 and 4. In each figure, the x-axis corresponds to the thresholds applied on the test collections (i.e. 0 to 100) and the y-axis to the value of the metric (i.e. precision, recall, and F-measure). The points in the graph represent the performance of the systems on the test set. Each system used optimised parameters tuned on the training set.

As shown in Figure 3 and 4, O-OCC is more accurate in extracting emotion from text for the F-measure and precision, across all emotions and thresholds, whereas Figure 5 shows that EmpathyBuddy is better in terms of recall. The reason for the lower recall of O-OCC compared to Empathy-

Buddy is (1) that O-OCC provides the dominant emotions in the sentence rather than probability values for the six Ekman emotions; and (2) O-OCC's output is mapped from the OCC model to the Ekman emotion model, which is error prone. Finally, Sysesketch has the lowest effectiveness in comparison to other systems in terms of precision, recall, and F-measure. This is probably due to the lack of accuracy of the Affective WordNet base list used in this system and/or the naive linguistic interpretation of the sentences.

O-OCC model is based on a deep linguistic analysis of a sentence. As such it has a better performance, it is important to understand what role different components will have on its effectiveness on large scale.

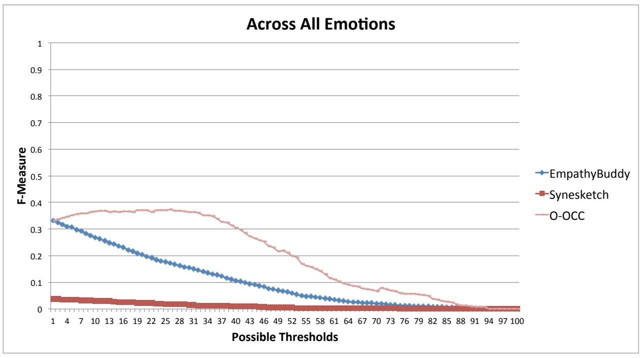

Figure 3: F-measure values across all emotions for all the thresholds on the news headline test set

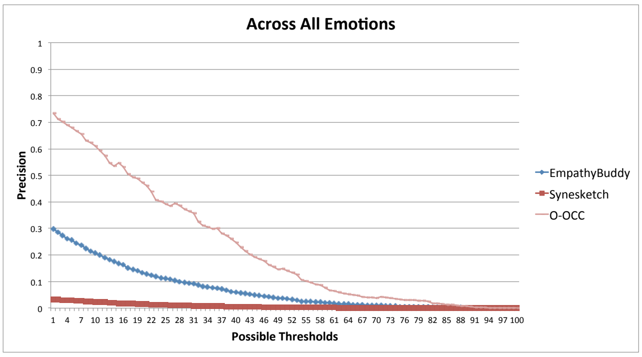

Figure 4: Precision values across all emotions for all the thresholds on the news headline test set

5. CONCLUSIONS

In this work, we presented a comparative study of emotion extraction techniques. The use of emotions in IR is hindered by lack of understanding on the effectiveness of the emotion extraction techniques. Therefore, in this paper we implemented an state-of-the-art emotion extraction method proposed in [3] (O-OCC) and compared its effectiveness with two open-source emotion extractor systems, i.e Empathy-Buddy and Synesketch. Our findings showed that O-OCC is more accurate in terms of precision and F-measure. In future work, we intend to study the effect of such methods in IR tasks such as browsing and retrieval, and also to study

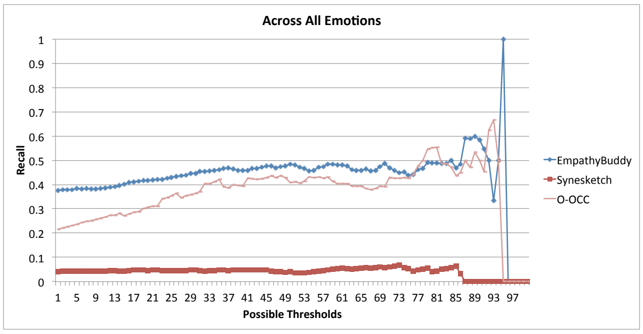

Figure 5: Recall values across all emotions for all the thresholds on the news headline test set

the effect of assigning emotion to a document based sentence level emotion analysis.

Acknowledgement This work was partially supported by the EU FP7 LiMoSINe project (288024).

6. REFERENCES

[1] Attardi, G., Simi, M.: Blog Mining through Opinionated Words. TREC'06) (2007)
[2] Moshfeghi, Y., Zuccon, G., Jose, J.M.: Using emotion to diversify document rankings. In: ICTIR. (2011) 337–341
[3] Masum, S.M.A., Prendinger, H., Ishizuka, M.: A Linguistic Interpretation of the OCC Emotion Model for Affect Sensing from Text. Affective Information Processing (2009) 45–73
[4] Izard, C.: The Psychology of Emotions. Plenum Pub Corp (1991)
[5] Ekman, P., Oster, H.: Facial Expressions of Emotion. Annual Reviews in Psychology **30**(1) (1979) 527–554
[6] Liu, H., Lieberman, H., Selker, T.: A model of textual affect sensing using real-world knowledge. IUI'03 (2003) 125–132
[7] Ortony, A., Clore, G., Collins, A.: The cognitive structure of emotions. Cambridge Univ Pr (1990)
[8] Liu, H., Selker, T., Lieberman, H.: Visualizing the affective structure of a text document. CHI'03 (2003) 740–741
[9] Strapparava, C., Valitutti, A.: WordNet-Affect: an affective extension of WordNet. In: LREC'04. (2004) 1083–1086
[10] Krcadinac, U.: Textual emotion recognition and creative visualization. Master's thesis, University of Belgrade (2008)
[11] Miller, G.: WordNet: a lexical database for English. Communications of the ACM **38**(11) (1995) 39–41
[12] De Marneffe, M., MacCartney, B., Manning, C.: Generating typed dependency parses from phrase structure parses. In: LREC'06. (2006)
[13] Masum, S.M.A., Prendinger, H., Ishizuka, M.: Sentiment assessment of text by analyzing linguistic features and contextual valence assignment. AAI'08 **22**(6) (2008) 558–601
[14] Strapparava, C., Mihalcea, R.: Semeval-2007 task 14: Affective text. In: Proceedings of the SemEval'07. (2007) 70–74