

EMOTIVE ONTOLOGY: EXTRACTING FINE-GRAINED EMOTIONS FROM TERSE, INFORMAL MESSAGES

Martin D. Sykora, Thomas W. Jackson, Ann O'Brien and Suzanne Elayan
*Department of Information Science, Loughborough University, Loughborough University,
Loughborough, LE11 3TU, United Kingdom*

ABSTRACT

With the uptake of social media, such as Facebook and Twitter, there is now a vast amount of new user generated content on a daily basis, much of it in the form of short, informal free-form text. Businesses, institutions, governments and law enforcement organisations are now actively seeking ways to monitor and more generally analyse public response to various events, products and services. Our primary aim in this project was the development of an approach for capturing a wide and comprehensive range of emotions from sparse, text based messages in social-media, such as Twitter, to help monitor emotional responses to events. Prior work has focused mostly on negative / positive sentiment classification tasks, and although numerous approaches employ highly elaborate and effective techniques with some success, the sentiment or emotion granularity is generally limiting and arguably not always most appropriate for real-world problems.

In this paper we employ an ontology engineering approach to the problem of fine-grained emotion detection in sparse messages. Messages are also processed using a custom NLP pipeline, which is appropriate for the sparse and informal nature of text encountered on micro-blogs. Our approach detects a range of eight high-level emotions; anger, confusion, disgust, fear, happiness, sadness, shame and surprise. We report f-measures (recall and precision) and compare our approach to two related approaches from recent literature.

KEYWORDS

Sparse Text Analysis, Ontologies, Sentiment Analysis, Emotion Analysis, Information Retrieval, Twitter.

1. INTRODUCTION

Recently an entire industry of commercial services has grown around the automated monitoring of public relationship related chatter on social-media (Tollinen et al. 2012)¹. The voluntary sector, such as the crisis mapping community (Tapia et al. 2011) as a rule, often gets involved on social-media, in terms of monitoring and organising relieve efforts during conflicts, natural or manmade disasters (Kumar et al. 2011), hence over recent years specific tools that help to automate the monitoring of social-media content have been developed by this community (e.g. Rogstadius et al. 2011). The government has also shown interest in similar systems (e.g. Preotiuc-Pietro et al. 2012). These kinds of tools are necessary in order to deal with the enormous amounts of social-media content being generated every day, as mostly manual / human based monitoring is not feasible. Instead, automated tools and techniques to monitor and help analyse social-media content are required (Johansson et al. 2012).

Our EMOTIVE project was predominantly concerned with measuring emotions and their expression within a wide-range of different types of sparse and informal messages. This can be useful in monitoring public reaction to events, i.e. event characterization, but also in event clustering or user-profiling. We leveraged prior theoretic work on emotions, from the fields of psychology, social psychology, and social-media systems to construct an ontology of emotions. Indeed at the core of our work is a fine-grained evaluation of emotions, since prior work has mostly focused on simplistic categories, such as negative / positive sentiment classification (Pang and Lee 2008, Liu 2010, Thelwall et al. 2012). Although numerous approaches employ highly elaborate and effective techniques with some success, the sentiment or emotion granularity is generally limiting; whereas EMOTIVE ontology was designed to detect a wide range of well recognised human emotions, such as ‘surprise’, ‘disgust’, or ‘confusion’, but at the same time differentiate emotions by strength (e.g. ‘uneasy’, ‘fearful’, ‘petrified’).

The remainder of this article is organized as follows. Section 2 introduces some background literature and the importance of emotion detection in social media streams is highlighted. Section 3 describes how the EMOTIVE ontology was constructed and some relevant technical details are also discussed. The Twitter specific NLP pipeline implementation and emotion matching process of ontology terms against the stream of sparse, informal messages is presented in section 4. Section 5 employs a golden (human annotated), test dataset in order to evaluate performance of our approach against two recent, state of the art approaches from literature. Finally, section 6 draws some conclusions and possible future work is discussed.

2. BACKGROUND

The research field of sentiment analysis has developed a variety of algorithms to automatically detect sentiment in text, and sentiment analysis of user generated content from social-media is now a well-established research area (Pang and Lee 2008, Liu 2010). Applications range from product and company related sentiment monitoring (Pang and Lee 2008), generic *real-time*

¹ This is corroborated by the emergence of relevant companies, such as; Attensity, Crimson Hexagon, Sysomos, Brandwatch, Vocus, Socialradar, Radian 6, Ecairn, Simplify360 and others.

“polling” on Twitter (Pak and Paroubek 2010) to gauging public emotional response to crises and national security concerns (Johansson et al. 2012) or terrorist attacks (Cheong and Lee 2011). However, the notions of affect and sentiment have been rather simplified in current state-of-the-art, often confined to their assumed overall polarity (i.e. positive / negative), with the POMS moods in Bollen et al. (2011), Lansdall-Welfare et al. (2012) and de Choudhury and Counts (2012) being the exceptions. Another problem with polarity-centric sentiment classifiers is that they generally encompass a vague notion of polarity that bundles together emotion, states and opinion. Unfortunately in the field of psychology, social psychology and affective computing there is still a lot of debate over human emotions and as Grassi (2009) points out, there is no common agreement about which features are the most relevant in the definition of an emotion and which are the relevant emotions and their names. In this work we solely consider emotions expressed in text content, and we specifically consider definitions of emotion taxonomies proposed by Ekman (1971), Plutchik (1980), Drummond (2004) and Izard (2009). Although all these authors arrived at somewhat different sets of human emotions, we look for their consensus and select a combined set of relevant emotions, see section 3 for more details.

As an attempt at a finer-grained emotions’ analysis, Bollen et al. (2011) employed a custom vocabulary that represents six states; tension, depression, anger, vigor, fatigue, and confusion. Their vocabulary extends the original POMS (Profile of Mood States) vocabulary, and contains 793 terms to improve matching of terms on sparse (Twitter) messages (793 instead of the original 65 terms, as they added synonyms from Wordnet and Roget’s New Millennium Thesaurus). However, only two of the POMS states are real emotions as tension, depression, vigor and fatigue represent an individuals’ states, rather than explicit and unambiguous emotions. For instance depression is a complex state that may be due to experiencing several emotions simultaneously, such as shame, sadness and confusion (Frijda 1986). Lansdall-Welfare et al. (2012) looked at the prevalence of the four emotions; anger, fear, joy and sadness, as they occur on Twitter during a variety of events. The terms (only single word expressions) associated with emotions were taken from the “Wordnet Affect” lexicon, stemmed and matched against tweets. de Choudhury and Counts (2012) employed a more thorough strategy than Lansdall-Welfare et al. in devising their emotionally charged terms. The authors used five established sources to develop a mood lexicon: 1-ANEW (Affective Norms for English Words), 2-LIWC (Linguistic Inquiry and Word Count), 3-EARL (Emotion Annotation and Representation Language), 4-basic emotions provided in Ortony and Turner (1990), and 5-list of moods provided by the blogging website LiveJournal. All these mood terms were evaluated by Amazon Mechanical Turk workers, as to whether the mood state actually represented a valid emotion in its own right, on a 1 – 7 Likert scale (*1 indicated not a mood at all, 7 meant definitely a mood*); only turkers having an approval rating greater than 95% were allowed, combining 12 different turkers’ ratings, a list of those words where the median rating was at least 4, and the standard deviation was less than or equal to 1 was constructed. The final set of mood words contained 203 terms, which were grouped under eleven basic headings; attentiveness, fatigue, fear, guilt, hostility, joviality, sadness, self-assurance, serenity, shyness and surprise. Emotion analysis (or advanced sentiment analysis) arguably provides a more fine-grained picture of affect than simple polarity sentiment analyses. A detailed profile of emotions in a piece of text has its uses in studies on information diffusion, event characterisation / clustering, and personality trait characterisation of users on social-media.

A recent and quite in-depth overview of prior academic work in the sentiment analysis field is provided in Thelwall et al. (2012), and the authors also discuss the three main approaches to sentiment detection; *machine learning*, *lexicon / linguistic analysis*, and *polarity estimation from term co-occurrence*. *Machine learning* approaches usually involve the training of a machine learning algorithm on human-annotated corpus, where features such as n-grams, parts-of-speech, etc. are used to train the classifier algorithms. *Lexicon* based approaches identify the presence of terms from a lexicon of known sentiment bearing words, and a further linguistic analysis of the context is usually performed to deal with negations of terms that further emphasise the matched sentiment words. This is also the approach that is employed in the EMOTIVE system described in this paper, except that we draw on emotion terms within an ontology with a far richer (and explicit) semantic representation than basic emotion term-lexicons. A state-of-the-art sentiment analysis method developed by Thelwall et al. (2012), SentiStrength-2, employs a lexicon and linguistic analysis approach, against which our technique is also evaluated in section 5. *Polarity estimation from term co-occurrence* method identifies the likely average polarity of words within pieces of text based on co-occurrence with respect to a set of seed words of unambiguous, known sentiment polarity. Turney (2002) introduced the idea of utilising web search engines to estimate relative co-occurrence frequencies for this purpose, and this is often used with such approach. Taboada et al. (2011) and Thelwall et al. (2012) both point out that although powerful, one drawback to non-lexicon based approaches is that they are likely to identify terms that associate with a specific sentiment but do not directly express it, such as *feel*, *Iraq*, and *late*. These techniques are more likely to learn domain specific, topical terms, typically associated with a given sentiment. This obscures detection of the direct expression of sentiment and might often not be desired.

In addition to detecting sentiment, many techniques attempt to predict a strength score of the expressed sentiment. For instance, SO-CAL (a lexicon based sentiment polarity analysis method), simply determines the final polarity score by computing average sentiment strengths of words in a piece of text (Taboada et al. 2011). SentiStrength-2 on the other hand outputs two integers; both in range 1 to 5, but one for positive sentiment and a separate score for negative sentiment. This is based on the assumption that a short piece of text (a tweet or status update) can contain both sentiment polarities at any one time (Thelwall et al. 2010). In our approach we compute a strength score for each distinct basic emotion; hence there can be eight separate scores in a piece of text, one for each emotion.

3. ONTOLOGY CONSTRUCTION

Our approach towards the problem of fine-grained emotions detection is based around semantic knowledge encoded within a custom built emotions ontology. The ontology covers eight basic emotions, negations, intensifiers, conjunctions, interjections, and contains information on the perceived strength (also known as activation level) of individual emotions, whether individual terms are slang or used in standard English and their associated POS (Parts-of-Speech) tags, where this aids to resolve ambiguity. Not only single words are represented but also multi-word phrases, which can further contain patterns, e.g. ‘*blew _blank_ _blank_ away*’ – ‘*blew me totally away*’, ‘*blew my mind away*’, and substrings such as prefixes or suffixes, e.g. the *[..]phobic* suffix, and similar expressions would match.

The ontology structure and in-depth study of language containing emotional expressions, required for the construction of the ontology, were performed by an English language and literature PhD level research associate, with training in linguistics and discourse analysis, during a three month time-period. In order to develop the ontology our RA has sifted through around 600MB of cleaned Tweets on 63 different UK-specific events / search-terms based Twitter datasets (collected using the Twitter *Search API*). Events as varied as the 2012 UK autumn floodings, Belfast riots, Olympic parade-games, Manchester PC shootings, the government reshuffle, and seasonal events such as Christmas 2012 and Diwali 2012, were amongst the events for which tweets were reviewed. This manual analysis focused on identifying commonly used explicit expressions of emotion.

The selection of basic emotions and their representative terms with strengths for each emotion were based on prior work – i.e. Izard (2009), Ekman (1971), Plutchik (1980), Drummond (2004) – and also on the commonly encountered emotions within Twitter messages themselves. Currently the ontology represents the following spectrum of eight basic emotions:

- | | |
|---------------------|--|
| 1. <i>Anger</i> | (e.g. <i>enraged, infuriated, peeved, in a tizzy...</i>) |
| 2. <i>Confusion</i> | (e.g. <i>chaotic, distracted, perplexed, confuzzled...</i>) |
| 3. <i>Disgust</i> | (e.g. <i>appalling, beastly, bullshit, scuzzy...</i>) |
| 4. <i>Fear</i> | (e.g. <i>cold feet, goose bumpy, petrified, scary...</i>) |
| 5. <i>Happiness</i> | (e.g. <i>blissful, chuffed, delighted, in high spirits...</i>) |
| 6. <i>Sadness</i> | (e.g. <i>depressed, devastating, duff, grief stricken...</i>) |
| 7. <i>Shame</i> | (e.g. <i>abashment, degrading, hang head in shame, scandalous...</i>) |
| 8. <i>Surprise</i> | (e.g. <i>astonished, disbelief, gobsmacked, off guard...</i>) |

These basic emotions are based on Ekman's emotions and are hence closely associated with the Plutchik's wheel of emotions (Plutchik 1980). All the eight basic emotions are represented except for '*anticipation*' and '*trust*', which were replaced with '*confusion*' and '*shame*', and the strengths of emotions also correlate with Plutchik's taxonomy. '*Confusion*' and '*shame*' were introduced on the basis of their use in other work and as these taxonomies differ for a variety of nuanced reasons, but were also introduced for the fact that we noticed a frequent expression of these emotions with their differing levels of strengths as reactions to a varied range of events on Twitter. Essentially we tried to consolidate these categories as we found they best relate to online expression of emotions. Table 3.1 highlights some of the basic emotional states that were identified and used in prior work. The greyed out emotions are not present in our ontology, and some other emotion classes were merged, such as Drummond's '*depression*' and '*hurt*'. All emotionally charged words were either added or excluded based on their contemporary use on Twitter.

EMOTIVE ONTOLOGY: EXTRACTING
FINE-GRAINED EMOTIONS FROM TERSE, INFORMAL MESSAGES

Table 1. Basic emotions identified by Drummond, Ekman Izard and Plutchik (grey colour indicates the emotion is not present in our ontology).

Drummond	Ekman	Izard	Plutchik
Anger	Anger	Anger	Anger
Caring	Disgust	Contempt	Anticipation
Depression	Fear	Disgust	Disgust
Fear	Happiness	Fear	Fear
Happiness	Sadness	Guilt	Joy
Hurt	Surprise	Interest	Sadness
Inadequateness		Joy	Surprise
Loneliness		Sadness	Trust
Remorse		Shame	
		Surprise	

In order to ensure a good recall of emotions, we also looked at OOV (Out of Vocabulary) terms of a large Twitter dataset in order to identify any commonly employed slang. Wordnet synset synonym lists of emotional expressions were also generated from our large Twitter datasets and manually reviewed. Other resources such as *Dictionary.com*, *Thesaurus.com* and the *Oxford English* online dictionary and the *Merriam-Webster* online dictionary were utilised in order to determine the strengths of the words' emotional intensity. Some of these dictionaries, however, did not include slang phrases popular among social website users and slang words or phrases that were endogenous within the relevant United Kingdom geographic areas (e.g. Leicestershire and Nottinghamshire areas). In order to detect as many slang expressions that convey emotions as possible, we examined websites such as *Urbandictionary.com*, *Internetslang.com* and other websites such as the *Leicestershire Slang Page*, the *Dictionary of Slang*, and the *Online Slang Dictionary* which also includes locations and origins of slang words². Emotional terms and activation levels identified and used in work by Choudhury et al. (2012) and the lexicon lists of intensifiers, negators and words of basic sentiment used in SentiStrength-2 by Thelwall et al. (2012) were also reviewed and integrated into our final ontology. This helped to ensure that as large a set of explicit emotions as possible was covered. One important consideration for our ontology was that for a term to be included in it, it had to be an explicit expression of emotion, rather than a term that expresses a personal state which may be vaguely associated with an emotion. This is an important consideration, since as opposed to other work we do not focus on mood states, but we consider direct expressions of emotions to be solely indicative.

Part-of-speech tagging was performed in order to facilitate any further shallow parsing of messages, but it is primarily employed in disambiguating emotion terms from the ontology, e.g. 'cross' can be an emotion indicating anger but it can also be a motion-verb, however when it is used as an adjective it most likely is an emotion (this type of knowledge is represented in the ontology).

² <http://dictionary.com>, <http://thesaurus.com>, <http://oxforddictionaries.com>, <http://merriam-webster.com>,
<http://urbandictionary.com>, <http://internetslang.com>, <http://www.angelfire.com/fl2/slang/>,
<http://www.peevish.co.uk/slang/>, <http://onlineslangdictionary.com>

3.1 Technical Details

The ontology was designed in Protégé using OWL and RDF, and the relevant code for efficiently loading up the ontology into memory was developed in Python; internally the eight basic emotions are represented as a set of custom Hashtable and Trie wrapper objects, together with details on valid intensifiers, conjunctions, negators, interjections, perceived emotional strengths and valid parts-of-speech tags with information on slang and formal use. The low level OWL interaction implementation was built with the widely used open-source Python library – namely, RDFLib. As a side note, the intermediate step of loading the ontology serialisation into a memory triple graph representation can be time consuming. Therefore our ontology was serialised in the NT serialisation format, as this was found to be most efficient, see table 2, below.

Table 2. RDFLib time taken (in seconds) to load an example ontology serialisation into memory (repeated 500 times). *Loading an XML serialisation takes nearly three times longer than parsing the NT serialisation.*

Format	Min	Avg	Max	Std. Dev.
XML	7.31	9.79	19.38	1.88
N3	3.74	4.98	9.51	1.05
NT	2.87	3.93	8.34	0.99

Stemming (using Porter and Lancaster stemmers), and Lemmatising individual emotional terms was considered, but deemed inappropriate for our needs, due to potential tendency to misrepresent and mismatch words. Instead we included specific term inflections directly into the ontology and automatically generated some basic aliases and dealt with common misspellings (e.g. ain't, aint, etc.). Including basic aliases and common misspellings increased the total term count threefold. Conjunctions, intensifiers and negators were ignored unless they occurred within the semantic context of an emotion. The semantic context was simply assumed to be bound by a sentence (*regex based sentence splitting was performed*), alternatively by a window of proximity related tokens. Further technical details, relating to natural language processing (NLP) are given in section 4.

4. THE MESSAGE PROCESSING PIPELINE

Natural language processing of formal and grammatically well written texts is a well-established research area (Manning and Schütze 1999); however, due to enforced brevity of messages (e.g. 140 characters or less on Twitter), textual content commonly encountered on social media, contains extensive use of slang, short-hand syntax, incorrect spelling, repeated letters, repeated words, inconsistent punctuation, odd Unicode glyphs, emoticons and overall a high proportion of OOV (Out-Of-Vocabulary) terms. Sentences are often not grammatically bound nor constructed properly, for instance tweets often start with a verb where the subject is implied and hashtags (and at-mentions) can be used as part of the tweet's message - e.g. *'rushing to the station, need 2get home in time for #dinnertime & avoid the #londonriots #hate it!!'*. This complicates text processing considerably and it was proposed that an entirely retrained NLP pipeline is hence necessary (Ritter et al. 2011). We performed our own evaluations of the NLP processing pipeline for sparse type of messages; comparing various

approaches from literature for tokenisation, text normalisation and parts-of-speech tagging. In our pipeline, messages were initially segmented into sentences based on punctuation. We employed a tweaked version of Potts (2011) regex based tokeniser, with O'Connor et al.'s (2010) emoticon matching rules, which cover a wide range of emoticons. The tokenisation itself was customised, to break up certain tokens on their suffixes or prefixes, for instance *lipophobic* would become *lipo + phobic*. This allows for fast and efficient matching of substrings using Hashtables; as was mentioned in section 3.1, the ontology was loaded into a set of Hashtable and Trie data-structures, which were used to match the ontology terms. A Trie (or prefix tree) was used to store phrases loaded from the ontology, where each edge in the Trie represents a word in a phrase. A Trie facilitates an efficient search for the longest possible phrase. The detection of emotions was done on a token per token basis, where each token in a tweet message was looked up in a Hashtable, if no match was found, then the token with the sequence of all remaining tokens in the message would be queried against the Trie, to see if a phrase emotion exists. Once an emotion was found, intensifiers, negators and conjunctions that occur before the emotion token would be matched too, if they satisfy some constraints. An intensifier is only picked up if it is actually related to the emotion; sentence boundaries or alternatively token proximity help to achieve this. Also, for instance a conjunction is only matched when it acts on two separate emotions (e.g. "Things were *bad* but thr so *fantastic* now!!"). Gimpel et al. (2010) proposed an intricate technique for Part-Of-Speech tagging; unfortunately the processing speed was about 10 tweets a second³, which simply does not scale-up to social media streams. Their POS tagger was mostly slowed down by separating proper nouns from other parts-of-speech; however, we did not need this separation to resolve ambiguous emotion matches. Instead we trained a Brill-tagger, with pre-tagged tokens from a tri-gram tagger, with backoff to bi-gram, uni-gram and a regex pattern tagger. This custom POS tagger is faster in the multiples and achieved a respectable 0.88 accuracy (baseline on tweet style messages is around 0.7, Ritter et al. 2011) on unseen Twitter datasets from Gimpel et al. (2010).

Finally as emotions, intensifiers, negators and conjunctions are matched we keep a running total of the emotional charge score for each found emotion. Once strength scores for each emotion are computed, an overall emotionality strength score can be given by taking the sum, average or maximum for each tweet. The high-level diagram in figure 1, below, coarsely illustrates the NLP pipeline.

³ The authors have since released a newer and optimised version (i.e. they replaced the CRF classifier), although the current versions' performance may still not be sufficient for some applications (Owoputi et al. 2012).

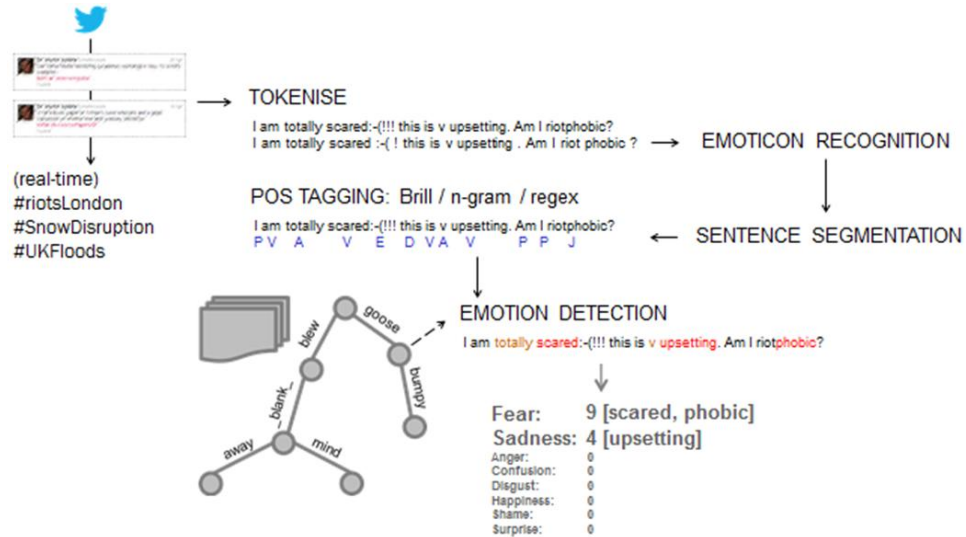


Figure 1. EMOTIVE’s NLP pipeline: incoming tweets; tokenisation; emoticon recognition; sentence segmentation; POS tagging; Hashtables/Tries of Emotions, Negators, etc. are matched against tweets.

5. EVALUATION

A golden-dataset of 150 tweets (312 entities), for evaluating the EMOTIVE ontology was annotated. The dataset is essentially a collection of tweets from a varied set of events (e.g. UK September 2012 floods, Manchester police shootings, etc. – *the annotated dataset is available at <http://emotive.lboro.ac.uk/resources/IJCSIS.html>*). These tweets are annotated with explicit expressions of emotions, and related negations, intensifiers, conjunctions, and interjections. Two RAs have independently annotated the dataset with a 99% agreement among the Tweets, where the remaining 1% was discussed between the RA’s and in all cases any disagreements resolved. In order to assess system performance; recall, precision and f-measure, were computed using an equivalent approach as used in *CoNLL-2003* shared task on NER (Tjong et al. 2003).

Table 3 summarises the performance of our ontology based matching, which highlights that the best f-measure (.962) was achieved when sentence segmentation, POS-tagging based ambiguity resolution of emotions and a maximum token neighbourhood rule for intensifiers, negators and conjunctions were employed. Recall being 1 indicates that the ontology covers all expressions and relevant terms that actually occur in the dataset; however, precision varies as false-positives arise due to various sources of ambiguity.

Table 3. Performance of ontology matching (ablation tests)

Matching	F-measure	Precision	Recall
Sent. Seg. + POS + Rules	.962	.927	1
Sent. Seg. + Rules	.952	.909	1
POS + Rules	.935	.879	1
Sent. Seg. + POS	.929	.868	1

EMOTIVE ONTOLOGY: EXTRACTING
FINE-GRAINED EMOTIONS FROM TERSE, INFORMAL MESSAGES

A major cause of false-positives is the incorrect detection of relevant intensifiers, see second row in table 3 for an example tweet of such miss-classifications. The third and fourth row in table 3, highlight how performance suffers when sentence segmentation and maximum token neighbourhood rule are left out from the processing pipeline, respectively. Both are effective for resolving false positive matches. The second row in table 3 highlights the improvement of matching ambiguous emotions, based on POS-tags (see section 4). It correctly detects terms such as ‘chicken’ and ‘cross’ that would otherwise be mismatched as valid emotions.

Table 4. Example system matched tweets (basic matching, i.e. without any rules / sentence segmentation)

Tweet	Comment
wow it is quite emotional , starting to well up here #olympicparade so proud of our athletes and our nation	‘wow’, ‘emotional’, ‘well up’, with so (int.) proud
i love indian holidays so much , like diwali , holi , onam and much more , they're so joyful ! i wanna go to india ! ? ? anyways , happy diwali !	4 incorrectly identified intensifiers (striked-out)

Given that there is considerable disagreement as to; what are valid emotions, actual emotive expressions, and the categories of basic emotion groups considered by other authors, it can be problematic to compare different emotion detection systems. In order to place our ontology matching approach in context of prior work on emotion detection in social-media streams, we evaluated Choudhury et al. (2012)’s 11 emotion category-based lexicon matching (see section 2). Only emotions were matched and evaluated, since Choudhury et al. did not incorporate intensifiers, negators and similar terms into their lexicon. Terms were also stemmed (using Porter-stemmer) to see whether recall could be improved. The achieved performance of matching the 203 emotion terms and stemmed version of the terms on the golden-dataset was; .744 / .668 (f-measure), .830 / .658 (precision) and .674 / .680 (recall), respectively. The removal of term inflections (using stemming) improved recall very slightly at the cost of mismatching many terms, due to higher matching likelihood. In any case our own system performs well in comparison to the .744 f-measure. Although our ontology contains 306 emotional expressions and phrases, hence the difference in recall.

Table 5. Ontology matched emotional expression strength score vs. SentiStrength-2 sentiment score (correlations)

Our Scoring	SentiStrength-2
Summed Score	.428 / .353
Avg. Score	.356 / .289
Max. Score	.396 / .325

*[Spearman’s rank correlation / Kendall’s rank correlation]
[All correlations were significant at p<.0005 (two-tailed test)]*

To gauge the performance of our emotion strength scoring approach, we evaluated our system against Thelwall et al. (2012)’s SentiStrength-2. Each emotion expression in our ontology has a score associated, and intensifiers, negators and conjunctions further modify this score. SentiStrength-2 considers these linguistic elements also, although for each message (i.e. Tweet) two polarity score sentiments are generated. Hence we computed SentiStrength’s absolute sentiment score value and compared it to the sum, avg. and max. score of our basic emotion types scores, for each tweet in the golden-dataset. Table 5 highlights a consistent and

statistically significant correlation; which indicates that we are measuring in line with a sentiment scoring state-of-the-art system. However, there are considerable differences, for instance the tweet in row one, in table 4, SentiStrength would have missed intensifiers ‘quite’, and ‘so’, and the emotional phrase ‘well-up’. Our system considers this tweet highly emotional giving it a score of 17 (8.5–avg. / 10–max.), whereas SentiStrength only scores it 3. More generally SentiStrength-2 simply does not score extremely emotional tweets as high as we tend to, as emotions are specifically and more fully represented in our system.

6. CONCLUSION

An ontology based system for fast and efficient capturing of a wider and a more comprehensive range of human emotions as opposed to existing systems, has been proposed in this paper. Several aspects of the emotions ontology construction and implementation details, specific to Twitter and related sparse informal text-based social media streams were presented. The emotion matching and matching of relevant linguistic elements that affect the expressed emotion, such as intensifiers, negators, conjunctions and interjections, were evaluated on a golden-dataset. The performance of our system was compared to another very recent lexicon based emotion matching approach from literature, and the emotion strength scores devised from the ontology matching process were put into perspective by comparing these scores to SentiStrength-2. Current approaches can be improved with better modeling techniques and techniques from classical NLP need to be tweaked further, to better deal with highly informal and sparse social-media datasets. In the future we intend to further extend the ontology and evaluate it extensively, on a much larger test-set of tweets from a variety of different events and against other systems.

REFERENCES

- Bollen J., Mao H. and Pepe A., 2011. Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Spain
- Cheong M. and Lee V. C. S., 2011. A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter, *Journal of Information Systems Frontiers – Springer* 13 (1), pp. 45-59
- Choudhury M. and Counts S., 2012. The Nature of Emotional Expression in Social Media: Measurement, Inference and Utility, Technical Report: Microsoft.
- Drummond T., 2004. Vocabulary of Emotions [Online], North Seattle Community College, [last viewed 9.1.2012]. Available from <http://www.sba.pdx.edu/faculty/mblake/448/FeelingsList.pdf>
- Ekman P., 1994. All emotions are basic. *The nature of emotion: Fundamental questions* 15-19.
- Frijda N., 1986. *The Emotions*, Cambridge University Press
- Gimpel K., Schneider N., O'Connor B., Das D., Mills D., Eisenstein J., Heilman M., Yogatama D., Flanigan J. and Smith N., 2010. Part-of-speech tagging for twitter: Annotation, features, and experiments, Technical Report.
- Grassi M., 2009. Developing HEO human emotions ontology, *Biometric ID Management and Multimodal Communication*, Springer Berlin Heidelberg, pp. 244-251

EMOTIVE ONTOLOGY: EXTRACTING
FINE-GRAINED EMOTIONS FROM TERSE, INFORMAL MESSAGES

- Izard C. E., 2009. Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology* **60**, 1-25.
- Johansson F., Brynielsson J. and Quijano M., 2012. Estimating Citizen Alertness in Crises Using Social Media Monitoring and Analysis, Proceedings of the European Intelligence and Security Informatics Conference (EISIC), Odense (Denmark)
- Kumar S., Barbier G., Abbasi M. A., Liu H., 2011. TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief, Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona (Spain)
- Lansdall-Welfare T., Lampos V. and Cristianini N., 2012. Effects of the Recession on Public Mood in the UK, Proceedings of the 21st international conference companion on World Wide Web, Lyon (France)
- Liu B., 2010 (2nd ed). *Sentiment Analysis and Subjectivity*, Handbook of Natural Language Processing, CRC Press - Taylor and Francis Group, USA
- Manning C. D. and Schütze H., 1999. *Foundations of statistical natural language processing*, MIT Press. USA
- Pak A., and Paroubek P., 2010. Twitter as a corpus for sentiment analysis and opinion mining, Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC2010), Malta
- Pang B. and Lee L., 2008. Opinion Mining and Sentiment Analysis, *Foundations and Trends in Information Retrieval* **2** (1-2), pp. 1-135
- Plutchik R., 1980. *Emotion: A Psychoevolutionary Synthesis*. Longman Higher Education.
- Preotiuc-Pietro D., Samangooei S., Cohn T., Gibbins N. and Niranjana M., 2012. Trendminer: An Architecture for Real Time Analysis of Social Media Text, Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, Dublin (Ireland)
- Potts C., 2011. Potts Twitter-aware Tokeniser - <http://sentiment.christopherpotts.net/code-data/happyfuntokenizing.py>, [last viewed 29.3.2013]
- Revelle W. and Scherer K., 2009. Personality and Emotion, *Oxford Companion to the Affective Sciences* (Oxford University Press), pp. 303-305
- Ritter A., Clark S., Mausam and Etzioni O., 2011. Named Entity Recognition in Tweets: An Experimental Study, Proceedings of Conference on Empirical Methods in Natural Language Processing, Edinburgh (UK)
- Rogstadius J., Kostakos V., Laredo J. and Vukovic M., 2011. A real-time social media aggregation tool: Reflections from five large-scale events, Proceedings of the European Conference on Computer-Supported Cooperative Work (ECSCW), Aarhus (Norway)
- O'Connor B., Krieger M. and Ahn D., 2010. TweetMotif: Exploratory Search and Topic Summarization for Twitter, Proceedings of the International AAAI Conference on Weblogs and Social Media, Washington DC (USA)
- Ortony A. and Turner T. J., 1990. What's basic about basic emotions? *Psychological Review* **97**, pp. 315-331
- Owoputi O, O'Connor B., Dyer C., Gimpel K., Schneider N., 2012. Part-of-Speech Tagging for Twitter: Word Clusters and Other Advances, Technical Report (School of Computer Science, Carnegie Mellon University)
- Taboada M., Brooke J., Tofiloski M., Voll K. and Stede M., 2011. Lexicon-based methods for sentiment analysis, *Computational Linguistics* **37** (2), pp. 267-307
- Tapia A., Bajpai K., Jansen B., Yen J. and Giles L., 2011. Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations, Proceedings of the 8th International ISCRAM Conference, Lisbon (Portugal)

- Thelwall M., Buckley K., Paltoglou G., Cai D. and Kappas A., 2010. Sentiment Strength Detection in Short Informal Text, *Journal of the American Society for Information Science and Technology* **61**, pp. 2544–2558
- Thelwall M., Buckley K. and Paltoglou G., 2012. Sentiment Strength Detection for the Social Web, *Journal of the American Society for Information Science and Technology* **63**, pp. 163-173
- Tjong E. F., Sang K., Meulder F. D., Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition, *Proceedings of the ACL seventh conference on Natural Language Learning*.
- Tollinen A., Jarvinen J. and Karjaluo H, 2012. Social Media Monitoring in the industrial Business to Business Sector, *World Journal* **2** (4), pp. 65-76