

Good News or Bad News: Using Affect Control Theory to Analyze Readers' Reaction Towards News Articles

Areej Ahothali

David R. Cheriton School
of Computer Science
University of Waterloo
Waterloo, Ontario, N2L3G1
aalhotha@cs.uwaterloo.ca

Jesse Hoey

David R. Cheriton School
of Computer Science
University of Waterloo
Waterloo, Ontario, N2L3G1
jhoey@cs.uwaterloo.ca

Abstract

This paper proposes a novel approach to sentiment analysis that leverages work in sociology on symbolic interactionism. The proposed approach uses Affect Control Theory (ACT) to analyze readers' sentiment towards factual (objective) content and towards its entities (subject and object). ACT is a theory of affective reasoning that uses empirically derived equations to predict the sentiments and emotions that arise from events. This theory relies on several large lexicons of words with affective ratings in a three-dimensional space of evaluation, potency, and activity (EPA). The equations and lexicons of ACT were evaluated on a newly collected news-headlines corpus. The ACT lexicon was expanded using a label propagation algorithm, resulting in 86,604 new words. The predicted emotions for each news headline was then computed using the augmented lexicon and ACT equations. The results had a precision of 82%, 79%, and 68% towards the event, the subject, and object, respectively. These results are significantly higher than those of standard sentiment analysis techniques.

1 Introduction

Natural language texts are often meant to express or impact individuals' emotions. Recognizing the underlying emotions expressed in or triggered by a text is essential to understanding the full meaning that a message conveys. Sentiment analysis (SA) researchers are increasingly interested in investigating natural language processing techniques as well

as emotion theories to identify sentiment expressions in natural language texts. They typically focus on analyzing subjective documents from the writer's perspective using frequency-based word representations and mapping them to categorical labels or sentiment polarity, in which text (e.g., sentence or document) is associated, respectively, with either a descriptive label or a point in a continuum.

In this paper, we focus on analyzing objective statements. Unlike subjective statements, that contain an explicit opinion or belief about an object or aspect (e.g., "*Seventh Son is a terrible movie*"), objective statements present only factual information (e.g., "*protestors were arrested in NYC*"). Despite the limited research on fact-based sentiment analysis, many factual statements may carry or evoke sentiments (e.g., news articles, blog comments, etc.). Thus, there is a need for a new approach that attaches sentiment scores to objective (factual) statements as well as to the components that make them up. For example, a sentence like "*x kills y*" will clearly evoke a negative sentiment for the reader, and highly negative (yet different) sentiments towards each x and y (angry about x and sorry about y). Further, our affective evaluation of each entity participating in the event might affect our judgement of the situation. For example, if we know that x is a victim and y is a criminal, then the triggered sentiment will be more positive in general, positive towards x , but still negative towards y .

The proposed method in this paper builds a contextual model that maps words to a multi-dimensional emotion space by using Affect Control Theory (ACT). ACT is a social psychological

theory of human social interaction (Heise, 2007). ACT proposes that peoples’ social perceptions, actions, and emotional experiences are governed by a psychological need to minimize deflections between culturally shared fundamental sentiments about social situations and transient impressions resulting from the dynamic behaviours of interactants in those situations. Each event in ACT is modeled as a triplet: actor, behavior, and object. Culturally shared “fundamental” sentiments about each of these elements are measured in three-dimensions: Evaluation, Potency, and Activity (EPA). The core idea of ACT is that each of the entities (actor, behaviour and object) participating in an event has a fundamental affective sentiment (EPA value) that is shared among members of a culture, and the combination of entities in the event generates a transient impression or feeling that might be different from the fundamental sentiment. Transient impressions evolve over time according to empirically measured temporal dynamics. Emotions are functions of the difference between fundamental sentiments and transient impressions. EPA profiles of concepts can be measured with the semantic differential (Osgood, 1957), a survey technique whereby respondents rate affective meanings of concepts on numerical scales with opposing adjectives at each end (e.g., {good, nice} ↔ {bad, awful} for E; {weak, little} ↔ {strong, big} for P; {calm, passive} ↔ {exciting, active} for A). Affect control theorists have compiled databases of a few thousand words along with average EPA ratings obtained from survey participants who are knowledgeable about their culture (Heise, 2010). For example, the culturally shared EPA for “*mother*” in Ontario, Canada, is [2.74, 2.04, 0.67], which is quite good, quite powerful, and slightly active. The “*daughter*” EPA is [2.18, -0.01, 1.92], which is quite good, but less powerful and more active than “*mother*”.

ACT is advantageous for sentiment analysis applications, especially those that aim to analyze descriptive events or factual content for the following reasons: (1) EPA space is thought to provide a universal representation of individuals’ sentiment; (2) the transient feeling accumulates over the course of several events and, hence, it can model larger structures in text; (3) impression formation can be computed towards different entities in the sentence,

giving a more fine-grained description of the sentiments; (4) the EPA values are empirically driven, reflecting a real evaluation of human participants; and (5) the interaction between terms in ACT is compatible with the linguistic principle of a compositional semantic model, which states that the meaning of a sentence is a function of its words (Frege, 1892).

Our method uses the impression and emotion formation equations of ACT to predict human sentiments towards events. We decompose each sentence into subject-verb-object and then associate subject with actor, verb with behaviour, and object with object in ACT. We then compute the predicted emotion towards each of actor, behaviour and object using ACT equations. We use a semi-supervised method based on Wordnet-similarities to assign EPA values to words that are not in ACT lexicons. We evaluated the viability of using ACT in sentiment analysis on a news headlines dataset that we collected and annotated. This approach yielded a precision between 71% and 82% on the news headline dataset. These results are significantly higher than those results from a method that was trained using bag-of-words and Sentiwordnet lexicon.

The rest of the paper is organized as follows: the first section presents the related work in emotions elicitation and sentiment analysis fields; the following two sections describe the proposed method, providing details about affect control theory, impression formation equations, and the lexicon induction method; the last section presents the dataset used in this paper, and discusses the results obtained using our proposed methods.

2 Related Work

Emotions have been studied extensively in disciplines like anthropology, psychology, sociology, and more recently in computer science. In recent years, fields like affective computing (AC) (Picard, 2000), human-computer interaction (HCI) (Brave and Nass, 2002), and sentiment analysis (Feldman, 2013) are also contributing to this area of research, by computationally modeling and evaluating existing theories. Most of the proposed methods in SA and AC have used the appraisal theory by representing emotions with discrete labels such as happy, sad, etc. (Ekman, 1992). Only several studies have adopted di-

mensional theories by representing emotions (core affects) in three- or two-dimensional space: arousal, valence, and sometimes dominance (Russell, 2003). Although psychological evidence show that words or any other events are emotionally perceived in more than one dimension (Barrett, 2006; Russell, 2003), multi-dimensional models have rarely been used in sentiment analysis.

In sentiment analysis research, two main machine learning (ML) methods are used: supervised learning or unsupervised learning methods. Supervised learning methods generally use the occurrence frequencies of words that indicate opinion/sentiment and then classify these occurrences as either positive or negative using one of the common classification methods (e.g., Maximum Entropy, support vector machine (SVM)). Many combinations of features have been tried, such as part-of-speech tagging, n-grams, term weighting, sentiment lexicons, term presence, and syntactic dependencies (Pang et al., 2002; Lin and He, 2009). In contrast, unsupervised learning approaches often determine the semantic orientation (SO) of a sentence or document by calculating the difference between the point-wise mutual information (PMI) (Church and Hanks, 1990) for its phrases or words with sentiment seeds (Turney and Littman, 2002; Turney, 2002).

Other approaches have built more-structured models, augmenting the standard bag-of-words technique with appraisal groups, which are represented as sets of attribute values in a semantic taxonomy (Whitelaw et al., 2005) or considering the interaction between words by using a tree structured CRF based on (Nakagawa et al., 2010). Another study has taken the interaction between words into account by utilizing the compositional formal semantic models based on Frege's principle (Frege, 1892). A framework is proposed in (Coecke et al., 2010) in which a sentence vector is a function of the Kronecker product of its words. This approach was evaluated on several datasets and showed promising results and improvement over non-compositional approaches (Grefenstette et al., 2010; Grefenstette and Sadrzadeh, 2011; Mitchell and Lapata, 2008).

Much sentiment analysis work has focused on extracting emotions from the writer's perspective; only several recent studies have tackled the problem of predicting readers' sentiments. Both (Lin et

al., 2008) and (Yang et al., 2009) used bag-of-words and linguistic features to classify Chinese news articles (Yahoo!-Kimo news) into one of the user emotional ratings (happy, sad, surprise, etc.). Another study presents a multi-label classification approach that uses words' polarity, and semantic frame features to classify news article into categorical emotions (Bhowmick, 2009). As a part of the SemEval-2007 task, a number of approaches have been proposed to binary-classify news headlines into positive or negative sentiment (Strapparava and Mihalcea, 2007; Katz et al., 2007; Chaumartin, 2007).

3 ACT for Sentiment Analysis

3.1 Background

Affect Control theory (ACT) is a new version of symbolic interactionism (Mead, 1938) created by David Heise (Heise, 1987). ACT proposes that individuals process gestures (including words or events) as symbols or concepts shared among groups of people or culture. These fundamental meanings or 'fundamental sentiments' are defined in a three-dimensional space of EPA profile. ACT also proposes that individuals try to maintain the transient impressions, which are generated from the interactions of the events' elements (subject-verb-object), close to the fundamental sentiments in the EPA space. ACT models emotions as arising because of the differences between fundamental sentiments and transient impressions. For example, a person who is powerful but is made to feel powerless will feel "angry". ACT equations (i.e., impression formation equations) are obtained through empirical studies, and they model the process by which the fundamental sentiments of elements in the events are combined to generate a transient impression.

The affective meanings in ACT consist of three components: Evaluation (good versus bad); Potency (powerful versus powerless); and Activity (lively versus inactive). Each affective meaning is measured on a scale from -4.3 (infinitely bad, powerless, or inactive) to +4.3 (infinitely good, powerful, or lively). These meanings are attached to concepts corresponding to identities, behaviors, settings, and modifiers. According to ACT, people from the same cultures and gender share the same fundamental sentiments (EPA) about world concepts (Berger and

Zelditch, 2002; Heise, 2007).

Lexicons of EPA values have been gathered for thousands of words from different languages and cultures including Germany, Japan, Canada and the USA. In general, within-cultural agreement about EPA meanings of social concepts is high even across subgroups of society, and cultural-average EPA ratings from as little as a few dozen survey participants have been shown to be extremely stable over extended periods of time (Heise, 2010). These findings may seem surprising in light of societal conflicts as evidenced, for instance, by competing political ideologies. Research has consistently shown that the number of contested concepts is small relative to the stable and consensual semantic structures that form the basis of our social interactions and shared cultural understanding (Heise, 2010).

3.2 Affect Control Theory

In ACT, each event has at least three elements: actor (subject, S), behavior (verb, V), and object (O). Each of these elements is represented by three values (EPA) that capture the fundamental sentiments they evoke in terms of evaluation, potency, and activity. The fundamental sentiment of an event according to ACT grammar is a nine-dimensional vector:

$$f = \{S_e S_p S_a V_e V_p V_a O_e O_p O_a\}$$

where e.g. S_e represents the fundamental sentiment about the subject (S) on the evaluation (e) dimension. The transient impression evoked by an event is another 9D vector:

$$\tau = \{S'_e S'_p S'_a V'_e V'_p V'_a O'_e O'_p O'_a\}$$

where fundamental EPA values are denoted by non-primed symbols and post-event EPA values are denoted by primed symbols. The transient impression τ is computed by multiplying t , a vector of features that are combinations of terms from the fundamental sentiment f , by a matrix M of prediction coefficients estimated by impression-formation research.

$$t = (1 \ S_e \ S_p \ S_a \ V_e \ V_p \ V_a \ O_e \ O_p \ O_a \\ S_e V_e \ S_e V_p \ S_e V_a \ S_p V_e \ S_p V_p \ S_p V_a \ S_a V_e \\ V_e O_e \ V_e O_p \ V_p O_e \ V_p O_p \ V_p O_a \ V_a O_e \ V_a O_p \\ S_e V_e O_e \ S_e V_p O_p \ S_p V_p O_p \ S_p V_p O_a \ S_a V_a O_a)$$

$$\tau = Mt \quad (1)$$

For example, the transient impression of the subject's valence, S'_e , using US male coefficients:

$$S'_e = -.98 + .48 S_e - .015 S_p - .015 S_a \\ + .425 V_e - .069 V_p - .106 V_a + .055 O_e \dots$$

This part of the equation shows that our evaluation of the subject/actor is affected mainly by how the valence of this person and action are perceived by others (positive large coefficients .48 and .425 for S_e and V_e). It also shows that powerful actors (subject) or behaviours (verb) are seen a bit negatively (negative coefficients -.015, -.069 for S_p and V_p).

We also can incorporate the location (settings), which indicates where the event took place such as school, country, and etc., can be achieved by adding the EPA values for the setting and use the coefficient values of the subject-verb-object-location (SVOL) grammar instead of subject-verb-object (SVO).

Modifiers in ACT are adjectives or attributes that modify actor or object (e.g. "good friend" or "abusive father"). The impression generated from combinations of identity with modifiers can be calculated as a linear combination of the EPA values of both the identity and modifiers.

$$c = B_1 p + B_2 i \quad (2)$$

where $p = \{P_e, P_p, P_a\}$, $i = \{I_e, I_p, I_a\}$, and $c = \{C_e, C_p, C_a\}$ are the EPA profiles for the modifier, identity, and the combination, respectively, and B_1 and B_2 are coefficients estimated from survey data. For example, the "father" affective rating is [1.84, 1.78, 0.02], "abusive" is [-2.23, 0.34, -0.02], and "abusive father" is [-1.51, 1.37, -0.21].

The deflection, which is defined as the discrepancy between the fundamental sentiment and the transient impression, is calculated by the squared Euclidean distance between the sentiments and impressions given the following equation.

$$d = \sum_i (f_i - \tau_i)^2 \quad (3)$$

The deflection does not indicate positive or negative emotions, but rather indicates whether or not the event met someone's expectation.

In ACT, the emotion triggered by an event is a function of the fundamental identity for actor or object, $i \equiv \{S_e, S_p, S_a\}$ or $i \equiv \{O_e, O_p, O_a\}$, and the transient identity for actor or object, $i' \equiv \{S'_e, S'_p, S'_a\}$ or $i' \equiv \{O'_e, O'_p, O'_a\}$. ACT uses the following equation to predict emotions, with empirically measured coefficients as follows:

$$\varepsilon \propto E (i' - I i - \delta) \quad (4)$$

where E is a 3×3 matrix coefficient of the emotion profile, I is a 3×3 matrix coefficient for identity, and δ is a vector of equation constants.

3.3 Emotion Elicitation Using ACT

We implement ACT to predict the triggered sentiment of a single sentence as follows: first we extract the subject, verb, object, setting, and modifiers (adjectives of subject and object) and look them up in the augmented ACT lexicon (see section 3.4) to get EPA values (fundamental sentiments, f) for each word (MacKinnon, 2006). We next compute the transient impression τ using f and Equations 1 and 2. After that, we compute the deflection using Equation 3, and the emotion towards the subject and object using Equation 4. We then map the resulting EPA scores for emotion to the nearest emotions label in ACT. The ACT dataset (MacKinnon, 2006) has 135 emotion labels, each with an EPA score (e.g., delighted= [2.04, 0.96, 1.48]), and we find the closest label using a Euclidean distance measure. We then compare these emotions ε towards subject and object to corresponding ground truth in the news headline dataset (see Section 4) using root mean squared error (RMSE) and mean absolute error (MAE). We also discretize the predicted ε and the ground truth into negative or positive EPA values, and compare accuracy of the discretized values.

To extract the sentence’s quintet (i.e., subject, verb, object, modifiers, and settings), we implement a search algorithm that takes a treebank parse tree (Socher et al., 2013) and returns a (subject, predicate, and object) triplet (Rusu et al., 2007). As Rusu *et al.*’s algorithm searches a parse tree of grammatically correct English sentences, we make some alterations to consider news headlines’ grammar. News headlines are often written to be short and precise (i.e., often not grammatically correct),

and usually consist of several noun phrases without articles or the verb “to be”. They are written in the present tense for current or past events (e.g, *terror strikes police base*). The past tense verbs are rarely used in news headlines, whereas passive voice sentences are common. The passive voice sentences are often written without an auxiliary verb, which make it hard for standard parsers to distinguish their verbs from past tense verbs and to extract the triplet accurately (e.g, *Six killed in accident*).

Our algorithm takes a parse tree and performs a breadth-first search and identifies the last noun descendent of the first noun phrase (NP) in the sentence as the subject and the previous descendent as the attributes (in Rusu *et al.*’s algorithm the subject is the first noun in (NP)). For example, in the sentence *Super Bowl-winning quarterback Russell Wilson divorces wife*, the actor/subject “Russell” is the last noun in the first noun phase and the previous adjectives and nouns are attributes (modifiers) of the subject. To locate the verbs, (similar to Rusu *et al.*’s algorithm) the algorithm searches the deepest verb phrase (VP) and returns the first verb (VB) descendent. If the verb is in the past tense, we transform it to passive voice. The algorithm (similar to Rusu *et al.*’s algorithm) returns the object that is co-located with the verb in the deepest verb phrase (VP). To extract the settings, we look for a noun phrase (NP) sub-tree that has a preposition (at, in, on) and return the last noun. This algorithm yields accuracies of 43%, 53% , and 26% with the ground truth (users’ annotations of the subject, verb, and object).

We also consider whether the verb type is transitive, which directly indicates positive or negative sentiment toward something (e.g., x killed y), or intransitive, which transfers sentiments into nouns (e.g., x provides help to y). For intransitive verbs, we choose the second verb as the behavior (verb) in the sentence (e.g., “ x provides help to y ” will be “ x helps y ”). We also used part-of-speech tagging to determine the elements of an event and to identify the places and names. The gender of the names is considered by training a naïve Bayes classifier on names-gender corpus of 5001 female and 2943 male names¹ which yielded an accuracy of 86% on classifying names according to their gender.

¹www.nltk.corpus

3.4 EPA lexicon Induction

The method described in the last section relies on a lexicon that maps words to EPA values. The ACT lexicon we used originally contains 2,293 words (original-EPA-lexicon) (MacKinnon, 2006). We augmented this lexicon by adding the Affective Norm for English Words (ANEW) (Bradley and Lang, 2010) data-set that contains 2,476 English words, and the Warriner *et al.* data-set (Warriner et al., 2013) that contains 13,915 words. ANEW and Warriner *et al.* data-sets were both scaled from the range of [1,9] to the range of [-4.3,4.3] using max-min scaling formula (Han, 2012). The min-max normalization performs a linear transformation for a given value x_i of A with a minimum and maximum value of $[min_A, max_A]$ to x'_i in range of $[min_B, max_B]$ given this formula:

$$x'_i = \frac{x_i - min_A}{max_A - min_A}(max_B - min_B) + min_B$$

Adding ANEW and Warriner lexicons generated lexicon of 17,347 words (extended-EPA-lexicon). We then randomly divided the extended-EPA-lexicon into a training-EPA-lexicon and a testing-EPA-lexicon, with 5,782 and 11,565 words, respectively. We used the training-EPA-lexicon to add more words, using a graph-based semi-supervised learning method called label-propagation, a technique that has been commonly used for NLP (Chen et al., 2006; Niu et al., 2005; Rao and Ravichandran, 2009; Zhu and Ghahramani, 2002; Blair-Goldensohn et al., 2008), and image annotation (Cao et al., 2008; Heckemann et al., 2006). Label-propagation is a transductive algorithm that propagates information from a set of labeled nodes (seed sets) to the rest of the graph through its edges (Zhu and Ghahramani, 2002). The label-propagation algorithm starts by adding all the synonyms and lemmas in WordNet of a specific part-of-speech (verb, noun, adjective, or adverbs) to the training-EPA-lexicon. This generates a set of labeled words $L = (X_l, Y_l)$ (the 5,782 words with EPA labels), and unlabeled words $U = (X_u, Y_u)$, from which we constructed undirected weighted graph $G = \{E, V, W\}$, where V is a set of vertices (all the words in the set), E is the weighted edges, and W is an $n \times n$ weight matrix (affinity matrix) n equal to the size vocabulary $|V|$.

We initialized the labeled nodes/words with the EPA value of the words observed in the training set, and the unlabeled nodes/words with zeroes. We computed the weight matrix using the WordNet similarity (Wu and Palmer, 1994) between the two words x_i and x_j . Wu and Palmer’s similarity is equal to the depth of the least common subsumer (LCS, the least common ancestor) divided by the summation of the depth of the two words in the WordNet taxonomy.

$$sim_{wup}(w_1, w_2) = \frac{2 * depth(LCS)}{depth(w_1) + depth(w_2)}$$

Each edge $E \in (v_i, v_j)$ has an associated weight T_{ij} , which is the row normalized weight w_{ij} of the edge between v_i and v_j . The labels are then propagated to adjacent nodes by computing $Y \leftarrow TY$. After each iteration the labeled nodes Y_l are reset to their initial values (see Algorithm 1).

Algorithm 1 ACT Label Propagation

procedure LABEL PROPAGATION(*Synests C*)
 Construct a Graph $G = \{V, E, W\}$
 Initialize T^0 and Y^0 matrices, $i \leftarrow 0$
repeat
 $Y^i \leftarrow TY^{i-1}$
 $Y_l \leftarrow L$
until Y converges
end procedure

The label propagation algorithm generated 167 adverbs, 3,809 adjectives, 11,531 verbs, and 81,347 nouns, with their EPA ratings in which 10,249 of them are in the testing-EPA-lexicon. To evaluate the validity of this approach, we compare our generated EPA ratings for these 10,249 words with those from the testing-EPA-lexicon (from ACT/ANEW/Warriner datasets). The results are shown in Tables 1 and 2. The resulting EPA are equally distributed between -4.3 and +4.3. We used two metrics to compare the results: root mean squared error (RMSE) and mean absolute error (MAE). These metrics were both close to 1.0 for E, P, and A, suggesting that there is a reasonable degree of agreement between the induced and manually labeled EPA values. It is worth mentioning that due to the limited numbers of adverbs in the ACT

lexicon, and because the Wordnet-similarity measure that compares only words of the same part-of-speech, only several adverbs were generated using label-propagation algorithm.

POS	W	RMSE			MSA		
		E	P	A	E	P	A
Adjectives	378	1.2	1.2	0.9	0.9	1.0	0.8
Adverbs	5	1.0	1.1	1.3	0.7	0.8	0.9
Verbs	2,787	1.2	1.1	0.8	1.0	0.9	0.6
Noun	7,079	1.3	1.1	0.9	1.0	0.9	0.7
Total	10,249	1.3	1.1	0.9	1.0	0.9	0.7

Table 1: The results of comparing the induced lexicon using label propagation and ground truth EPA values (POS= part-of-speech, W= the number of the induced words, MAS=mean absolute error, and RMSE= root mean squared error)

4 Datasets

We evaluated the proposed method on a newly collected news-headlines dataset. We collected 2080 news-headlines from a group of news websites and archives (BBC, CNN, Reuters, The Telegraph, Times, etc). The news headlines were selected randomly from the period from 1999 to 2014. Through Mechanical Turk, we recruited participants located in North America with more than 500 approved hits and an approved rate above 90%. We asked the participants to locate the subject (actor), behavior (verb), and object of each sentence and to indicate their emotions towards them and towards the event (as a whole) in the EPA format $\in [-4.3, +4.3]$ (where -4.3 indicates strongly negative EPA value and +4.3 indicates strongly positive EPA value). The dataset was annotated by at least three judges per headline. We excluded any ratings that were filled with blanks, zeros, or similar values in all the fields. We also excluded the answers that did not have the appropriate subject, verb, or object form (e.g., behavior=Obama, subject=as). We also excluded all the answers rated by less than three participants. This screening resulted in 1658 headlines that had a mean EPA rating of 0.80, 1.04, and 1.02. Of these, 995 headlines had a positive evaluation score and 663 headlines had a negative evaluation score. Some examples from this dataset can be seen in Table 5.

Words	Testing-EPA-lexicon	LP-lexicon
Incapable (adj.)	[-1.83, -1.40, -0.54]	[-1.56, -1.18, -2.59]
Wrongly (adv.)	[-1.96, -0.22, 0.17]	[-2.02, -0.23, 0.18]
Gauge (v.)	[0.12, -0.55, 0.13]	[0.18, -1.61, 0.25]
Loser (n.)	[-1.30, -1.75, 0.30]	[-1.14, -1.52, 0.28]

Table 2: Words and their EPA ratings from Testing-EPA-lexicon and LP-lexicon=label propagation lexicon

5 Results

Our model (the augmented lexicon, and ACT equations) was evaluated in predicting the evoked sentiment towards the headlines as a whole by comparing the discretized evaluation (E) score $\in \{0, 1\}$ (where 0/1 indicates negative/positive emotions, resp.) of the generated EPA to the ground truth. This evaluation was performed using different configurations (Table 3): (1) Using users’ annotated triplet (ACT-UA) (i.e., subject, verb and object), the model yielded a precision of 75% compared to the ground truth. (2) Using the parse tree triplet (ACT-PTT) (see Section 3.3 for details), the precision dropped to 71%. (3) Adding the adjectives (modifiers) and settings to the subject, verb and object, which we will called parse tree quintet (ACT-PTQ) yielded a higher precision, with 82% precision in comparison with the corresponding ground truth. These results were also compared to the results obtained from a standard sentiment classifier (STD-calssifier) that uses occurrence frequencies of positive vs. negative words using SentiWordNet (Das and Bandyopadhyay, 2010). This classifier yielded a precision of 57% in comparison to the ground truth (Table 3).

The parse tree quintet (ACT-PTQ) were also used to evaluate the ACT predicted emotions towards the actor (subject) and the object in the headline (Table 4). Three metrics were used in this evaluation: precision, MAS, and RMSE. We used precision to compare the discitized EPA scores $\in \{0, 1\}$ and MAS and RMSE to compare the real EPA scores $\in [-4.3, +4.3]$. As shown in Table 4, the RMSE and MAS are almost all less than 1.5, and the precision varies from 67% to 79% across discretized E,P,A for subject and object. To put these results in context, a difference of 1.4 in the EPA space would equate to the difference between “accusing” someone ($\{-1.03, 0.26, 0.29\}$) and “pun-

Classifier	Precision	Recall	F1-score
ACT-PTQ	82	67	73
ACT-UA	75	62	67
ACT-PTT	71	63	66
STD-classifier	57	51	53

Table 3: Results for sentiment classification of news headlines dataset using ACT and standard sentiment classification method, ACT-PTQ = ACT using the parsing tree quintet, ACT-UA = ACT using users’ annotation, ACT-PTT= ACT using the parsing tree triplet, STD-classifier= Standard classifier using SentiWordNet

Emotions	Precision			MAS			RMSE		
	E	P	A	E	P	A	E	P	A
ETS	79	74	76	1.11	1.25	1.27	1.38	1.24	1.33
ETO	67	68	67	1.09	1.26	1.27	1.33	1.56	1.54

Table 4: Comparison of predicted emotions towards the subject and object with the ground truth. ETS/ETO = Emotions towards subject/object, MAS=mean absolute error, and RMSE= root mean square error

ishing” someone ($\{0.19, 0.79, 0.76\}$), or between the identity of ”mother” ($\{2.48, 1.96, 1.15\}$) and ”girl”($\{1.96, 0.67, 0.99\}$), or between the emotion of ”joyful” ($\{2.43, 1.97, 1.33\}$) and ”euphoric” ($\{1.42, 1.09, 0.99\}$). These words seem quite close in an affective sense, which indicates that our sentiment analysis method is able to uncover sentiments at a level that is reasonable on an intuitive level, and shows the method’s power in uncovering sentiments about specific elements of sentences.

6 Discussion and Future Work

Human emotions are more complicated than several labels or binary scores. ACT models emotions in a three-dimensional space which is found to be a comprehensive and universal representation of human emotions, and models emotions towards event- or fact-based sentences as a combination of several emotions towards their entities (subject and object). To evaluate the effectiveness of using ACT in sentiment analysis, we chose to analyze news headlines as they represent real-world statements that describe single or multiple events/facts.

Analysing sentiment in news headlines is a challenging task for several reasons: (1) news headlines

are ungrammatically structured, which makes it hard for standard parsers to extract their words’ part-of-speech and dependency correctly; (2) they are written to be short and precise, providing little information for typical bag-of-word classifiers to work properly; (3) they are objective, containing words that might not exist in the commonly used sentiment lexicon. To overcome the limitation of news headline sentiment analysis, three main contributions that we present in this paper: (1) we extracted the sentences’ triplets by considering the headline grammar and structure; (2) we augmented ACT lexicon using label-propagation and word similarity; (3) we model the interaction between the words in the sentence by modelling the transient and fundamental sentiments.

The label propagation algorithm generated 96,853 words in which 10,249 of them are in the testing data set. The EPA values of these words were found to be quite close in the affective space to their corresponding ground truth, Table 1. Table 2 also shows several good examples of the generated EPA and their corresponding ground truth. The label-propagation results could be further improved by adding words antonyms and by employing another similarity measure. The results of predicting the sentiment towards the event and their entities as shown in Tables 4, 3, and 5 are computed using *only* the ACT lexicons and the ACT impression formation equations. With such a simple, parsimonious and theoretically well-grounded approach, we are able to compute fine-grained sentiment analysis in a dimensional space that is known to be a universal representation of human affect. Mapping these three-dimensional EPA scores to a specific emotion provides a detailed label for objects and subjects within a sentence, as shown in Table 5. For example, in a sentence like “*Russia says 4 militants killed in Dagestan siege*”, the reader will be feeling negative, yet different emotions towards the subject “*furious*” and the object “*sorry*”.

Table 5 (obtained with ACT-PTQ) shows the deflection (d), emotions towards the events (ε), and towards the subject and object (e_a) and (e_o) of some of the examples in the data set. As shown in Table 5, we can see that the deflection (d) is very high when we do not expect an event to occur (e.g., “*Baby dies after being left in car for over 8 hours*”). The deflection in this sentence is high (17.42) because the EPA value of the object “*Baby*” is equal to

Headline	d	ETS	ETO	T_e	ε	e_s	e_o
Press sees hope in Mecca talks	2.57	happy	reverent	1.33	2.50	1.53	1.59
Brazil deploys troops to secure borders for World Cup	2.32	proud	apathetic	1.7	1.61	0.66	1.22
Gunfire injures three Napoli fans	6.80	furious	melancholy	-1.13	-0.86	-1.25	0.54
Three political candidates slain before Iraqi vote	11.24	furious	sorry	-1.33	-1.46	-1.80	0.05
Lily Allen wins web music award	2.74	proud	reverent	1.67	2.45	1.46	1.36
Finland Air crash kills skydivers	12.54	furious	cheerless	-1.33	-3.20	-3.0	-0.10
Bomb kills 18 on military bus in Iran	3.40	impatient	overwhelmed	-1.6	-1.23	-2.3	-0.11
Russia says 4 militants killed in Dagestan siege	11.37	furious	sorry	-0.2	-1.46	-2.34	-0.58
Baby dies after being left in car for over 8 hours	17.42	furious	overwhelmed	-1.67	-2.10	-1.57	0.06
Female astronaut sets record	0.79	contented	reverent	3.50	0.91	1.37	0.46

Table 5: ACT model’s results on news headlines, d =deflection, ETS, ETO= emotion towards subject and object, T_e = emotions towards the event (ground truth), ε =emotions towards the event (ACT), and e_s, e_o = the evaluation value of the emotion towards subject and object. Parse elements are coded as: *subject*, *verb*, *object*, and *setting*.

[2.40, -2.28, 2.58], which is considered to be quite good, quite weak, and quite active identity and a car is considered to be a quite positive place, with EPA value equal to [1.62, 1.65, 2.01]. While if an event took place in a war zone or if the subject has negative evaluation, the deflection will not be very high (e.g., “Bomb kills 18 on military bus in Iran”) the deflection is equal to 3.40 because “Bomb” has a negative evaluation. In Table 5, we can see the emotions (ε) and the ground truth evaluation toward the events (T_e) are often quite close.

The aforementioned results are obtained by extracting single subject, verb, object, modifier, and setting. These results could further improved by taking adverbs (e.g., “lived happily”), phrasal verbs (e.g., “get along” and “get back”), numbers (e.g., “45 killed”), and negations (e.g., “no more funding”) into consideration. Finally, accumulating the emotions of multiple consequent behaviors could also be very useful. For example, in the sentence “Man arrested after beating cops in a restaurant” the behavior will be “arrested”, and “beating” is not taken into account. We could address this using more complex parse trees and accumulating the emotions of multiple behaviors by considering the previously generated sentiment as the fundamental sentiment. Finally, using ACT predictions to bootstrap supervised learning could also yield improvements.

7 Conclusions

We have proposed a new direction for sentiment analysis, employing Affect Control Theory (ACT) to assign different emotions towards events-

based/objective statements and their entities (subject, object). Unlike the majority of sentiment analysis models that are trained on highly subjective words to obtain descriptive labels, our model incorporates ACT, that models emotions as points in three-dimensional space, and analyzes how objective texts trigger different emotions. We use a semi-supervised method based on Wordnet similarities to compute emotional ratings for words not in the ACT lexicons. Evaluated on a news headline dataset, our model yielded higher accuracy than a widely used classifier, with a highest precision of 82%. We also analyzed the sentiment evaluation of ACT on (actor/subject and the object) in the news headlines, yielding a precision of 79% and 68% when analyzing the emotions towards the subject and the object, respectively. These results have been obtained without performing any supervised learning and without taking consequent behaviors, phrasal verbs, or sentence negations into account. Thus, they demonstrate the potential of ACT for sentiment analysis. Affect control theory can also handle consequent behaviors and modifiers. In future, we plan to augment our method with more complex levels of detail, gather more extensive datasets, and evaluate ACT for more precise and detailed sentiments.

Acknowledgments

The authors thank Tobias Schröder and Dan Lizotte for their thoughtful discussion and suggestions. Areej Alhouthali acknowledges the sponsorship of King Abdul Aziz University, Jeddah, Saudi Arabia. Jesse Hoey is supported in part by the Natural Sciences and Engineering Council of Canada (NSERC).

References

- Lisa Feldman Barrett. 2006. Are emotions natural kinds? *Perspectives on psychological science*, 1(1):28–58.
- Joseph Berger and Morris Zelditch. 2002. *New Directions in Contemporary Sociological Theories*. Rowman & Littlefield.
- Plaban Kumar Bhowmick. 2009. Reader perspective emotion analysis in text through ensemble based multi-label classification framework. *Computer and Information Science*, 2(4):P64.
- Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A Reis, and Jeff Reynar. 2008. Building a sentiment summarizer for local service reviews. In *WWW Workshop on NLP in the Information Explosion Era*, volume 14.
- MM Bradley and PJ Lang. 2010. Affective norms for english words (anew): Affective ratings of words and instruction manual (technical report c-2).
- Scott Brave and Clifford Nass. 2002. Emotion in human-computer interaction. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, pages 81–96.
- Liangliang Cao, Jiebo Luo, and Thomas S Huang. 2008. Annotating photo collections by label propagation according to multiple similarity cues. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 121–130. ACM.
- François-Régis Chaumartin. 2007. Upar7: A knowledge-based system for headline sentiment tagging. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 422–425. Association for Computational Linguistics.
- Jinxu Chen, Donghong Ji, Chew Lim Tan, and Zhengyu Niu. 2006. Relation extraction using label propagation based semi-supervised learning. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 129–136. Association for Computational Linguistics.
- Kenneth Ward Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29.
- Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *arXiv preprint arXiv:1003.4394*.
- Amitava Das and Sivaji Bandyopadhyay. 2010. Sentiword-net for bangla. *Knowledge Sharing Event-4: Task, 2*.
- Paul Ekman. 1992. Are there basic emotions?
- Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89.
- Gottlob Frege. 1892. On sense and reference. *Ludlow (1997)*, pages 563–584.
- Edward Grefenstette and Mehrnoosh Sadrzadeh. 2011. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404. Association for Computational Linguistics.
- Edward Grefenstette, Mehrnoosh Sadrzadeh, Stephen Clark, Bob Coecke, and Stephen Pulman. 2010. Concrete sentence spaces for compositional distributional models of meaning. *arXiv preprint arXiv:1101.0309*.
- Jiawei Han. 2012. *Data mining : concepts and techniques*. Elsevier/Morgan Kaufmann, Amsterdam Boston.
- Rolf A Heckemann, Joseph V Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. 2006. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126.
- David R Heise. 1987. Affect control theory: Concepts and model. *Journal of Mathematical Sociology*, 13(1-2):1–33.
- David R Heise. 2007. *Expressive order: Confirming sentiments in social actions*. Springer.
- David R. Heise. 2010. *Surveying Cultures: Discovering Shared Conceptions and Sentiments*. Wiley.
- Phil Katz, Matthew Singleton, and Richard Wicentowski. 2007. Swat-mp: the semeval-2007 systems for task 5 and task 14. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 308–313. Association for Computational Linguistics.
- Chenghua Lin and Yulan He. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 375–384. ACM.
- KH-Y Lin, Changhua Yang, and Hsin-Hsi Chen. 2008. Emotion classification of online news articles from the reader’s perspective. In *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT’08. IEEE/WIC/ACM International Conference on*, volume 1, pages 220–226. IEEE.
- Neil J. MacKinnon. 2006. Mean affective ratings of 2, 294 concepts by guelph university undergraduates, ontario, canada. In *2001-3 [Computer file]*.
- George Herbert Mead. 1938. *The philosophy of the act*, volume 3. Univ of Chicago Pr.
- Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. In *ACL*, pages 236–244.
- Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi. 2010. Dependency tree-based sentiment classification using crfs with hidden variables. In *Human Language*

- Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 786–794. Association for Computational Linguistics.
- Zheng-Yu Niu, Dong-Hong Ji, and Chew Lim Tan. 2005. Word sense disambiguation using label propagation based semi-supervised learning. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 395–402. Association for Computational Linguistics.
- Charles Egerton Osgood. 1957. *The measurement of meaning*, volume 47. University of Illinois Press.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.
- Rosalind W. Picard. 2000. *Affective computing*. MIT press.
- Delip Rao and Deepak Ravichandran. 2009. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 675–682. Association for Computational Linguistics.
- James A Russell. 2003. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.
- Delia Rusu, Lorand Dali, Blaz Fortuna, Marko Grobelnik, and Dunja Mladenic. 2007. Triplet extraction from sentences. In *Proceedings of the 10th International Multiconference "Information Society-IS*, pages 8–12.
- Richard Socher, John Bauer, Christopher D Manning, and Andrew Y Ng. 2013. Parsing with compositional vector grammars. In *In Proceedings of the ACL conference*. Citeseer.
- Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics.
- Peter Turney and Michael L Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus.
- Peter D Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics.
- Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207.
- Casey Whitelaw, Navendu Garg, and Shlomo Argamon. 2005. Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 625–631. ACM.
- Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 133–138. Association for Computational Linguistics.
- Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. 2009. Writer meets reader: Emotion analysis of social media from both the writer’s and reader’s perspectives. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 287–290. IEEE Computer Society.
- Xiaojin Zhu and Zoubin Ghahramani. 2002. Learning from labeled and unlabeled data with label propagation. Technical report, Technical Report CMU-CALD-02-107, Carnegie Mellon University.