# Classification and Pattern Discovery of Mood in Weblogs

Thin Nguyen, Dinh Phung, Brett Adams, Truyen Tran, and Svetha Venkatesh

Curtin University of Technology
thin.nguyen@postgrad.curtin.edu.au,
{d.phung,b.adams,t.tran2,s.venkatesh}@curtin.edu.au

**Abstract.** Automatic data-driven analysis of mood from text is an emerging problem with many potential applications. Unlike generic text categorization, mood classification based on textual features is complicated by various factors, including its context- and user-sensitive nature. We present a comprehensive study of different feature selection schemes in machine learning for the problem of mood classification in weblogs. Notably, we introduce the novel use of a feature set based on the affective norms for English words (ANEW) lexicon studied in psychology. This feature set has the advantage of being computationally efficient while maintaining accuracy comparable to other state-of-the-art feature sets experimented with. In addition, we present results of data-driven clustering on a dataset of over 17 million blog posts with mood groundtruth. Our analysis reveals an interesting, and readily interpreted, structure to the linguistic expression of emotion, one that comprises valuable empirical evidence in support of existing psychological models of emotion, and in particular the dipoles pleasure–displeasure and activation–deactivation.

## 1 Introduction

Mood is a state of the mind such as being happy, sad or angry. It is a complex cognitive process which has received extensive research effort, and debate, among psychologists about its nature and structure [9,10,6]. But better scientific understanding of what constitutes a 'mood' has ramifications beyond psychology alone: for neuroscientists, it might offer insight into the functioning of the human brain; for medical professionals working in the domain of mental health, it might enable better monitoring and intervention for individuals and communities.

Research like that cited above aims to understand psychological drives and structures behind human mental states, and typically does so with expensive methodologies involving questionnaires or interviews that limit the number of participants. By contrast, our work aims to classify and cluster mood based on pre-existing content generated by users, which is collected unobtrusively – a sub-problem known as *mood analysis* in sentiment analysis [8]. Text-based mood classification and clustering, as a sub-problem of opinion and sentiment mining, have many potential applications identified in [8].

However, text-based mood analysis poses additional challenges beyond standard text categorization and clustering. The complex cognitive processes of mood formulation make it dependent on the specific social context of the user, their idiosyncratic associations of mood and vocabulary, syntax and style which reflect on language usage, or the specific genre of the text. In the case of weblogs, these challenges are highlighted by bloggers in the expression of diverse styles, relatively short text length, and informal language, such as jargon, abbreviations, and grammatical errors. This leads us to investigate whether machine learning-based feature selection methods for general text classification are still effective for blog text. Feature selection methods available in machine learning are often computationally expensive, relying on labeled data to learn discriminative features; but the blogosphere is vast (reaching almost 130 million[1]) and continuing to grow, making desirable a feature set that works without requiring supervised feature training to classify mood. To this end, we turn our attention to the result of a study that intersects psychology and linguistics known as affective norm for English words (ANEW) [1], and propose its use for mood classification.

In addition to classification, clustering mood into patterns is also an important task as it might provide vital clues about human emotion structure and has implications for sentiment-aware applications. While the structure of mood organization has been investigated from a psychological perspective for some time [9], to our knowledge, it has not been investigated from a data-driven and computational point of view. We provide an analysis of mood patterns using an unsupervised clustering approach and a dataset of more than 17 millions blog posts manually groundtruthed with users' moods.

Our contribution is twofold. First, we provide a comparative study of machine learning-based text feature selection for the specific problem of mood classification, elucidating insights into what can be transferred from a generic text categorization problem for mood classification. We then formulate a novel use of a psychology-inspired set of features for mood classification which does not require supervised feature learning, and is thus very useful for large-scale mood classification. Second, we provide empirical results for mood organization in the blogosphere on the largest dataset with mood groundtruth available today. To our knowledge, we are the first to consider the problem of data-driven mood pattern discovery at this scale.

The rest of this paper is organized as follows. Related work to feature selection methods for classification and clustering tasks in general text and in sentiment analysis is presented in Section 2. Work related to emotion measures in psychology is also examined in this section. Next, we present machine learning based feature selection schemes, together with the proposed ANEW feature set and linguistic analysis, applied to mood classification in two large datasets in Section 3. Section 4 presents the results of mood pattern discovery using an unsupervised learning technique, and is followed by some concluding remarks.

---

[1] From the state of the blogosphere 2008 at `http://technorati.com`

## 2   Related Background

For generic text categorization, a wide range of feature selection methods in machine learning has been studied. Most noticeably, Yang and Pedersen [14] conduct a comparative study on different feature selection schemes including information gain ($IG$), mutual information ($MI$), and $\chi^2$ statistic ($CHI$).

Other than term-class interaction based, another approach for selecting features is to consider term statistics. Thresholds for term frequency ($TF$) or document frequency ($DF$) are commonly used in feature reduction in data mining. A joint from these two, the term frequency–inverse document frequency ($TF.IDF$) scheme, is also popular in text mining which often outperform $TF$ and $DF$. Some works perform feature searching in narrower sets than over the entire vocabulary such as linguistic groups like the parts of speech ($POS$).

Related work making use of emotion bearing lexicon for sentiment analysis includes [2], where Dodds and Danforth use the valence values of ANEW [1] for estimating happiness levels in song lyrics, blogs, and the State of the Union.

Work related to mood clustering includes [5], where Leshed and Kaye group blog posts based on their moods to find mood synonymy.

Generally, emotions have been represented in dimensional and discrete perspectives. In the first methodology, emotion states are coded as combinations of some factors like valence and arousal. In contrast, the latter argues that each emotion has unique coincidence of experience, psychology, and behaviour [6]. We base our work on the dimensional mode for estimating the emotion sphere in blogosphere. Specifically, we use the circumplex model of affect [9,10] since it conceptualizes emotion states simply via valence and arousal dimensions, which can be computed using ANEW.

## 3   Textual-Based Mood Classification

### 3.1   Feature Selection Methods

Denote by $\mathcal{B}$ the corpus of all blogposts and by $\mathcal{M} = \{$sad, happy, ...$\}$ the set of all mood categories. In a standard feature selection setting, each blogpost $d \in \mathcal{B}$ is also labeled with a mood category $l_d \in \mathcal{M}$ and the objective is to extract from $d$ a feature vector $\boldsymbol{x}^{(d)}$ being as discriminative as possible for $d$ to be classified as $l_d$. For example, if we further denote by $\mathcal{V} = \{v_1, \ldots, v_{|V|}\}$ the set of all terms, then the feature vector $\boldsymbol{x}^{(d)} = [\ldots, \boldsymbol{x}_i^{(d)}, \ldots]$ might take a simple counting with its $i$-component $\boldsymbol{x}_i^{(d)}$ represents the number of times the term $v_i$ appears in blogpost $d$, a scheme widely known as bag-of-word representation.

**Term-based selection.** These are features derived with respect to a term $v$. Two common features are term and document frequencies where *term frequency* $TF(v, d)$ represents the number of times the term $v$ appears in document $d$, whereas *document frequency* $DF(v)$ is the number of blogposts containing the term $v$. It is also well-known in text mining that $TF.IDF(v, d)$ weighting scheme

can potentially improve discriminative power where $TF.IDF(v, d) = TF(v, d) \times IDF(v)$ with $IDF(v) = |\mathcal{B}|/DF(v)$ is the inverse document frequency. In this work, a term $v$ will be selected if it has high $DF(v)$ value, or high average values of $TF(v, d)$ or $TF.IDF(v, d)$ across all documents $d$ over a threshold.

**Term-Class interaction-based selection.** The essence of these methods is to capture the dependence between terms and corresponding class labels during the feature selection process. Three common selection methods falling into this category are information gain $IG(v)$, mutual information $MI(v, l)$ and $\chi^2$-statistics $CHI(v, l)$[14]. $IG(v)$ captures the information gain (measured in bits) when a term $v$ is present or absent; $MI(v, l)$ measures the mutual information between a term $v$ and a class label $l$; and lastly $CHI(v, l)$ measures the dependence between a term and a class label by comparing against one degree of freedom $\chi^2$ distribution.

**Affective Norms for English Words ($ANEW$).** Apart from feature sets learned from data, for sentiment analysis, some emotion bearing lexicons have been subjectively chosen by labor power could help. Among them is ANEW [1], a set of 1034 sentiment conveying English words. These words are rated in terms of valence, arousal, and dominance they could convey. We apply the proposed set of 1034 words in ANEW exclusively as the feature vector, which means each blogpost is represented as a sparse counting vector for these ANEW words.

### 3.2    Mood Classification Results

It has been shown that linguistic components such as specific use of adverbs, adjectives or verbs can be a strong indicator for mood inference [8]. Therefore, in this paper, we further run a part-of-speech tagger to identify all terms that can be tagged as verbs, adjectives and adverbs. The tagger used is the SS-Tagger [12] ported to the Antelope NLP framework, giving a reasonable accuracy[2]. Three term weighting-based ($TF$, $DF$, $TF.IDF$) and three term-class interaction-based ($IG$, $MI$, $CHI$) selection methods are employed in this experiment. These feature selection methods shall be applied either on all terms (unigrams) or with respect to a subset of terms tagged with a specific $POS$.

Our experimental design is to compare and contrast which feature selection methods work best and to examine the effect of specific linguistic components in the context of mood classification. For classification methods, we have experimented with many off-the-shelf classifiers such as SVM, IBK, C4.5 and so on, however, the naive Bayes classifier (NBC) consistently outperforms these methods and therefore we shall only report the results w.r.t NBC. For each run, we use ten-fold cross-validation and repeat 10 runs and report the average result. To evaluate the results, we report two commonly-used measures: *accuracy* and *F-score* (which is measured based on *recall* and *precision*).

**Effect of feature selection schemes and linguistic components.** We use two datasets, namely WSM09 and IR05, for the task of mood classification. The first, WSM09, is provided by Spinn3r as the benchmark dataset for ICWSM

---

[2] `www.proxem.com`

2009 conference[3] which contains 44 millions blogposts crawled between August and October 2008. We extract a subset from this dataset consisting only blogposts from LiveJournal and query LiveJournal to obtain the mood groundtruth entered by the user when the posts were composed. We only consider the moods predefined by LiveJournal and discard others, resulting in approximately 600,000 blogposts. To validate the generalization of a feature selection scheme, we also run it on another dataset (IR05) created in [7] which contains 535,844 posts tagged the predefined moods. To make comparison with previous results in [11], we examine three popular moods {sad, happy, angry} in this experiment. The full set of 132 mood categories will be reported in the next section.

We run the experiment over combination of feature selection methods on different linguistic subsets and report the top ten best results in Table 1.

**Table 1.** Mood classification results for different feature selection schemes and for different part-of-speech subsets. Different combinations of feature selection methods and *POS* subsets are run, but we report only the top ten results sorted in ascending order of *F-score*.

| WSM09 | | | | IR05 | | | |
|---|---|---|---|---|---|---|---|
| Selection method | Linguistic subsets | Accuracy | F-score | Selection method | Linguistic subsets | Accuracy | F-score |
| | ANEW | 0.713 | **0.697** | IG | Adjective | 0.738 | 0.709 |
| IG | Verb | 0.714 | 0.7 | | ANEW | 0.734 | **0.712** |
| TF.IDF | unigram | 0.744 | 0.738 | TF.IDF | AdjVbAdv | 0.759 | 0.749 |
| DF | AdjVbAdv | 0.75 | 0.745 | TF | AdjVbAdv | 0.76 | 0.75 |
| TF | AdjVbAdv | 0.751 | 0.745 | DF | AdjVbAdv | 0.76 | 0.75 |
| TF.IDF | AdjVbAdv | 0.754 | 0.748 | TF.IDF | unigram | 0.765 | 0.756 |
| DF | unigram | 0.753 | 0.752 | DF | unigram | 0.765 | 0.762 |
| TF | unigram | 0.753 | 0.752 | TF | unigram | 0.765 | 0.762 |
| IG | AdjVbAdv | 0.762 | 0.756 | IG | AdjVbAdv | 0.773 | 0.763 |
| IG | unigram | 0.776 | **0.774** | IG | unigram | 0.791 | **0.788** |

With respect to feature selection scheme, information gain (*IG*) is observed to be the best selection scheme. Other term-class interaction based methods do not perform well, noticeably mutual information (*MI*) does not appear in any of the top ten results. These observations are consistent with what reported in [14] for text categorization problem. However, different with conclusions in [14], we found that *CHI* performs badly for mood classification task and does not appear in any top ten results. Surprisingly, both *TF* and *DF* performs better than *TF.IDF* in all-term (unigram) cases, which otherwise has been known oppositely in text mining that *IF.IDF* is often superior although much more computationally expensive. Thus, *TF* or *DF* should be the alternative candidates for *IG* for the trade-off of computational cost.

The performance of feature selection schemes experimented is also agreeable well across two datasets as can be seen in Table 1, except for the first few
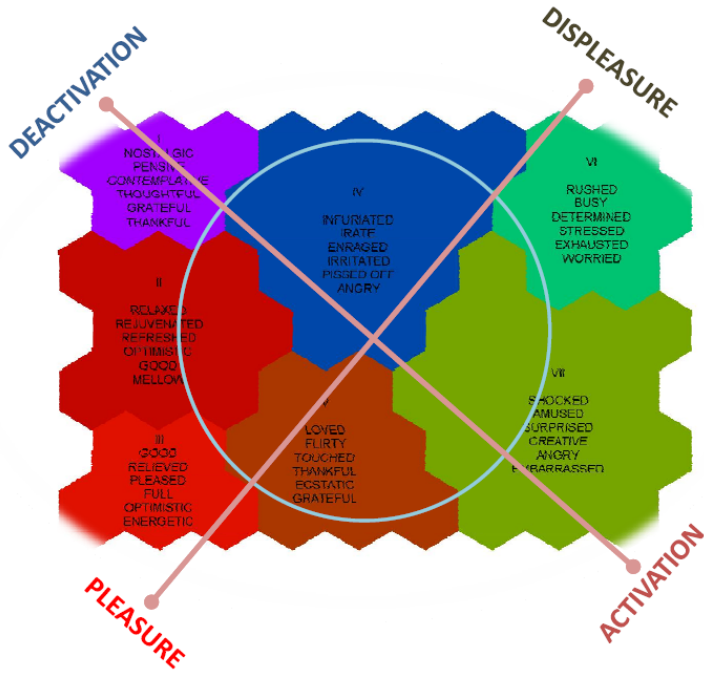
---

[3] http://www.icwsm.org/2009/data/

**Fig. 1.** Discovered mood structure map. Each cluster is annotated with the top six mood categories (best viewed in colour).

rows. Our best result stands at 77.4% *F-score* for WSM09 and 78.8% for IR05, which is higher than what reported in [11] (66.1%). With respect to the effect of linguistic components (which are not experimented in [11] and [5]), a combination of adjectives, verbs and adverbs (AdjVbAdv) dominates the top ten results and gives a very close performance to using all terms; noticeably using verbs or adjectives alone shows a good performance.

**Performance of ANEW.** Without the need of supervised feature selection stage, the result of ANEW feature is found to be very encouraging, appears in both top ten results across two datasets. The results across two datasets are also consistent, stand at approximately 70% *F-score* (still better than the best result reported in [11]).

## 4   Mood Pattern Discovery

While most of existing work has focused on supervised classification of mood, we are interested in discovering intrinsic patterns in mood structure using unsupervised learning approaches. Using a large, groundtruthed dataset of more than 17 millions posts introduced in [5], we aim to seek empirical evidences to answer various questions which have often posed in psychological studies. For example,

does mood follow a continuum in its transition from 'pleasure' to 'displeasure', or from 'activation' to 'deactivation'? Is 'excited' closer to 'aroused' or 'happy'? Does 'depressed' transit to 'calm' before reaching 'happy'?

We use a total of 132 predefined moods defined by livejournal.com[4] for the clustering task. Given a corpus of more than 17 millions posts, it means that feature selection schemes presented in section **3.1** are very expensive to perform; for example, computing $MI(v, l)$ for each pair (term, mood label) will take $O(|\mathcal{M}| \times |\mathcal{V}|)$ where $|\mathcal{M}| = 132$ (number of moods) and $|\mathcal{V}|$ is the number of unique terms which could be in the order of hundreds of thousands. Since our results in section **3.2** have shown that the proposed ANEW feature set gives comparable results, marginally lower ($\sim$8%) in the classification compared to the best result but can totally avoid the expensive feature selection step, we shall employ ANEW as the feature vector in this section.

We choose multidimensional scaling, in particular, self-organizing map (SOM) [3] for clustering purpose. We use the SOM-PAK package [4] and the SOM Toolbox for Matlab [13] to train and visualize the map. For training, an $9 \times 7$ map is used which accounts for nearly a half of the mood classes. Using the recommendations in [3], the horizontal axis is roughly 1.3 that of the vertical axis; the node topology is hexagonal, and the number of training steps is 32,000 (about 500 times of the number of nodes).

Due to space restriction, we omit coarse-level results and present in the Figure 1 the structures of the clusters discovered in which top six moods in each cluster are included.

Several interesting patterns emerge from this analysis. At the highest level, one can observe the general transition of mood from an extreme of *pleasure* (clusters II, III, and V) to *displeasure* (clusters IV, VI, VII). On the *pleasure* polar we observe the moods having very high valence values[5] such as *good* (7.47), *loved* (8.64) or *relaxed* (7), whereas on the *displeasure* end, we observe the moods having low valence values such as *enraged* (2.46) or *stressed* (2.33). Certain mood transition is also evidential, for example the cluster path IV-II-III presents a transition pattern from *infuriated* to *relaxed* and then to *good*. Though not strongly emerging as in the case of *pleasure* $\leftrightarrow$ *displeasure,* a global pattern of *activation* $\leftrightarrow$ *deactivation* is also observed based on the analysis of the arousal measure as shown in Figure 1. Our results are indeed favorable of the core affect model for human emotion structure studied in psychology [9,10], generally agreeable with the global mood structure proposed in there.

## 5   Conclusion

We addressed the problem of mood classification and pattern discovery in weblogs. While the problem of machine learning based feature selection for text categorization has been intensively investigated, little work is found for textual based mood classification which is often more challenging. Our first contribution

---

[4] These moods can be viewed at `http://www.livejournal.com`

[5] Measured based on a study on ANEW reported in [1].

is a comprehensive comparison of different selection schemes across two large datasets. In addition, we propose a novel use of ANEW features which do not require a supervised selection phase, and thus, can be applied for mood analysis at a much larger scale. Our results have recalled similar findings in previous results, but also brought to light discoveries peculiar to the problem of mood classification. Our newly proposed feature set has also performed comparatively well at a fraction of the computational cost of supervised schemes, and was further validated by the results of an unsupervised clustering exercise, which clustered 17 million blog posts, and provided a unique view of mood patterns in the blogosphere. In particular, this study manifests global patterns of mood organization that are analogous to the pleasure–displeasure and activation–deactivation dimensions proposed independently in the psychology literature, such as the core affect model for the structure of human emotion. This data-driven organization of mood could be of interest to a wide range of practitioners in the humanities, and has many potential uses in sentiment-aware applications.

# References

1. Bradley, M.M., Lang, P.J.: Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida (1999)
2. Dodds, P.S., Danforth, C.M.: Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. Journal of Happiness Studies, 1–16 (2009)
3. Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg (2001)
4. Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J.: SOM PAK: The self-organizing map program package. Technical report, Helsinki University of Technology (1996)
5. Leshed, G., Kaye, J.J.: Understanding how bloggers feel: recognizing affect in blog posts. In: Proc. of ACM Conf. on Human Factors in Computing Systems, CHI (2006)
6. Mauss, I.B., Robinson, M.D.: Measures of emotion: A review. Cognition & emotion 23(2), 209–237 (2009)
7. Mishne, G.: Experiments with mood classification in blog posts. In: Proc. of ACM Workshop on Stylistic Analysis of Text for Information Access, SIGIR (2005)
8. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), 1–135 (2008)
9. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
10. Russell, J.A.: Emotion, core affect, and psychological construction. Cognition & Emotion 23(7), 1259–1283 (2009)
11. Sara, S., Lucy, V.: Sentisearch: Exploring mood on the web. In: Proc. of Workshop on Weblogs and Social Media, ICWSM (2009)
12. Tsuruoka, Y.: Bidirectional inference with the easiest-first strategy for tagging sequence data. In: Proc. of ACL Conf. on HLT/EMNLP, pp. 467–474 (2005)
13. Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J.: SOM toolbox for Matlab. Technical report, Helsinki University of Technology (2000)
14. Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: Proc. of ICML, pp. 412–420 (1997)