

CruzAffect at AffCon 2019 Shared Task: A feature-rich approach to characterize happiness

Jiaqi Wu, Ryan Compton, Geetanjali Rakshit, Marilyn Walker, Pranav Anand, and
Steve Whittaker

UC Santa Cruz, 1156 High Street, Santa Cruz 95064, California
{jwu64, rcompton, grakshit, mawalker, panand, swhittak}@ucsc.edu

Abstract. We present our system, *CruzAffect*, for the CL-Aff Shared Task 2019. Cruz-Affect consists of several types of robust and efficient models for affective classification tasks. We utilize both traditional classifiers, such as XGBoosted Forest, as well as a deep learning Convolutional Neural Networks (CNN) classifier. We explore rich feature sets such as syntactic features, emotional features, and profile features, and utilize several sentiment lexicons, to discover essential indicators of social involvement and control that a subject might exercise in their happy moments, as described in textual snippets from the HappyDB database. The data comes with a labeled set (10K), and a larger unlabeled set (70K). We therefore use supervised methods on the 10K dataset, and a bootstrapped semi-supervised approach for the 70K. We evaluate these models for binary classification of *agency* and *social* labels (Task 1), as well as multi-class prediction for *concepts* labels (Task 2). We obtain promising results on the held-out data, suggesting that the proposed feature sets effectively represent the data for affective classification tasks. We also build *concepts* models that discover general themes recurring in happy moments. Our results indicate that generic characteristics are shared between the classes of *agency*, *social* and *concepts*, suggesting it should be possible to build general models for affective classification tasks.

Keywords: affective classification · well being theory · social connections.

1 Introduction

The overall goal of the CL-Aff Shared Task [4] is to understand what makes people happy, and the factors contributing towards such happy moments. Related work has centered around understanding and building lexicons that focus on emotional expressions [5, 9], while Reed *et al.* [7] learn lexico-functional linguistic patterns as reliable predictors for first-person affect, and constructed a First-Person Sentiment Corpus of positive and negative first-person sentences from blog journal entries. Wu *et al.* [12] propose a synthetic categorization of different sources for well-being and happiness targeting the private micro-blogs in Echo, where users rate their daily events from 1 to 9. These work aim to identify specific compositional semantics that characterize the sentiment of events, and attempt to model happiness at a higher level of generalization, however finding generic characteristics for modeling well-being remains challenging. In this paper, we aim to find generic characteristics shared between different affective

classification tasks. Our approach is to compare state-of-the-art methods for linguistic modeling to prior lexicons’ predictive power. While this body of work is broader in scope than the goals we are trying to address, they do include annotated sets of words associated with happiness as well as additional categories of psychological significance.

The aim of this work is to address the two tasks that are part of the CL-Aff Shared Task. The data provided for this task comes from the HappyDB dataset [1]. Task 1 focuses on binary prediction of two different labels, *social* and *agency*. The intention is to understand the context surrounding happy moments and potentially find factors associated with these two labels. Task 2 is fairly open-ended, leaving it to the participant’s imagination to model happiness and derive insights from their models. Here, we predict the *concepts* label using multi-class classification. We explore various approaches to determine which models work best to characterize contextual aspects of happy moments. Though the predictions of *agency* and *social* sound simpler than *concepts*, we expect that the best models for *agency* and *social* prediction could generate similarly optimal performance for *concepts*, assuming that the classes of *social*, *agency*, and *concepts* share common characteristics. To validate our assumptions, we build different models for general affective classification tasks and then try to gain a deeper understanding of the characteristics of happy moments by interpreting such models with the Riloff’s Autoslog linguistic-pattern learner [8, 12].

2 Agency and Social Classification

This work utilizes a bootstrapping approach to conduct semi-supervised learning experiments. This involves a three-step procedure: (1) train a model on the labeled data; (2) use the trained model to make predictions on the unlabeled data; and (3) train a new model using the combination of the labeled data and the predictions on the unlabeled data. Training each model involves a 10-fold cross-validation to evaluate the performance, while guaranteeing that the test set for each fold consists of gold-standard hand-labelled instances.

2.1 Feature Extraction

We explore different features to find those most informative for the prediction task. We aim to understand how syntactic features and emotional features compare to word embeddings, and whether the profile features improve the prediction results.

Syntactic Features: Our syntactic features are limited to *Part of Speech (POS)* tagging, by applying a POS tagger to count the relative frequencies of syntactic nouns, verbs, adjectives and adverbs, use of questions as well as tense and aspect information [10]. There are 36 POS features.

Emotional Features: We use 4 different types of emotional features. *LIWC v2007* [9] is a lexicon providing frequency counts of words indexing important psychological constructs, as well as relevant topics (Leisure, Work). The *Emotion Lexicon (EmoLex)* [5] contains 14,182 words classified into 10 emotional categories: Anger, Anticipation, Disgust, Fear, Joy, Negative, Positive, Sadness, Surprise, and Trust. The *Subjectivity Lexicon* is part of OpinionFinder [11]. It consists of 8222 stemmed and unstemmed

words, annotated by a group of trained annotators as either strongly or weakly subjective. Our last feature is our own regression model from prior work on predicting the *level of factual and emotion language*. Details about this model can be found in [3]. There are 94 features in total.

Word Embedding: We utilize *GloVe* [6] 100 dimension word vectors for word representation. GloVe is expected to encode distributional aspects of meaning.

Profile Features: The corpus include demographic features collected via a survey: *age, country, gender, married, parenthood, reflection, and duration*. To reduce sparsity, we convert the *country* feature into *language* feature, assuming that the people who speak the same language might share similar culture, and thus similar happy moments. After this conversion, we have 48 different languages from the 70 countries, the largest group is English (80.3%), then Hindi (15.9%), corresponding to 79.8% examples from USA and 15.87% examples from IND. We also bin age into groups, assuming that different age groups would have different general happy moments. The age groups, illustrated in Table 1, include kid (*age* < 10), teenager (*age* < 18), youth (*age* < 24), young adult (*age* < 40), middle age (*age* < 65) and elderly (*age* >= 65). There are 70 features after the feature preprocessing. We aim to test whether the features extracted from text are sufficient for affective sentiment analysis, and whether the profile features improve performance.

Table 1: The distribution of the age groups, and the probability of P(*agency*=yes) and P(*social*=yes) for each age group. The overall probability of P(*agency*=yes) is 0.74 and P(*social*=yes) is 0.53. The middle-age group is less likely to identify their happy moment with *agency* but more likely to identify the moment with *social*, while kids are more likely to identify their happy moment with both *agency* and *social*.

age	frequency	P(<i>agency</i> =yes)	P(<i>social</i> =yes)
elderly (<i>age</i> >= 65)	136	0.71	0.55
middle-age (<i>age</i> < 65)	1756	0.68	0.57
adult (<i>age</i> < 40)	7242	0.74	0.53
young (<i>age</i> < 24)	1401	0.79	0.50
teenager (<i>age</i> < 18)	0	0	0
kid (<i>age</i> < 10)	14	0.86	0.79
NaN	11	0.55	0.82

2.2 Classification Models

In this section, we compare the performance of traditional machine learning methods, such as Logistic Regression and XGBoosted Random Forest [2] and a Convolutional Neural Network (CNN) model [13].

Supervised Learning. For modeling the profile features, we apply logistic regression with a liblinear solver and balanced class weights. For modeling the syntactic features

and emotional features, we use XGBoosted Random Forest with out-of-the-box values for all parameters except for the following: 250 number of estimators, a learning rate of 0.05, and a maximum tree depth of 6. For the CNN model with word embedding, we explore its performance with different parameter settings. The best hyperparameters of the CNN model include filter size 3, multiple region size (2, 3, 4) and max pooling size 1, or filter size 4, multiple region size (2, 3, 4, 5) and max pooling size 1. The region size implies a windows size for N-grams. After getting the best hyperparameters, we train the model with word embeddings, and word embeddings concatenated with syntactic and emotional features, to test whether syntactic and emotional features improve performance. Figure 1 illustrates the CNN model with region size (2, 3, 4).

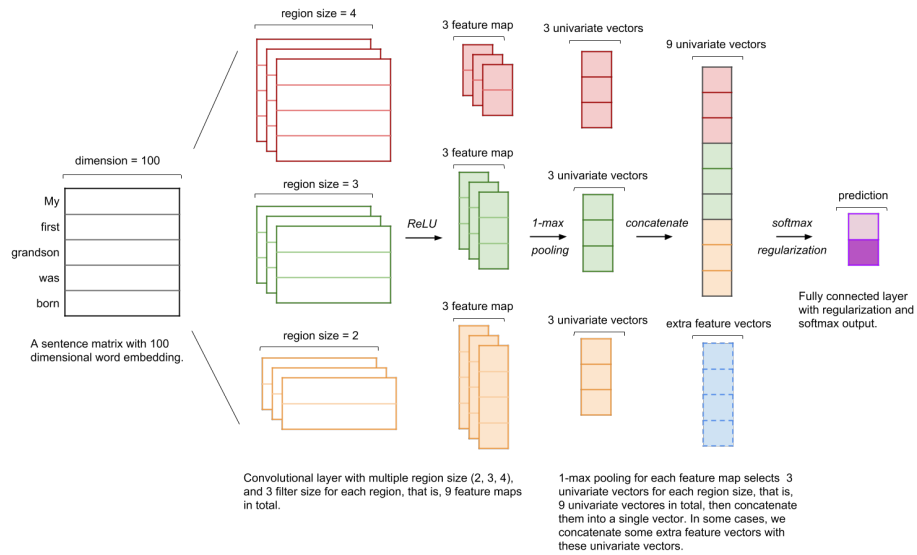


Fig. 1: A Diagram for the CNN model with region size (2, 3, 4) and filter size 3 for a single sentence. This architecture is cited from [13]. We have tuned some hyperparameters for our tasks. In some cases, we concatenate extra feature vectors, such as the syntactic features and the emotional features that are extracted from a sentence to the univariate vectors, then forward it to the softmax layer for the output.

Table 2 shows the 10-fold cross-validation results for the 10,560 labeled data. The macro-f1 score is reported along with the Precision, Recall, and Accuracy per label type. The Logistic Regression model, with profile features, yields F1-score 0.53 for *agency* prediction and 0.56 for *social* prediction. The CNN with word embedding outperforms Logistic Regression with F1-score 0.80 for *agency* prediction and 0.90 for *social* prediction. These results demonstrate that the happy moment contains enough information for the affective classification without profile features. The XGBoosted Forest with syntactic and emotional features also reaches a competitive F1-score of 0.78 for *agency* prediction and 0.90 for *social* prediction, meaning that these features are as represen-

Table 2: 10-fold cross-validation supervised learning for *agency* and *social* prediction.

Model	Features	Agency				Social			
		P	R	F1	Acc	P	R	F1	Acc
Logistic Regression	Profile	0.55	0.57	0.53	0.57	0.56	0.56	0.55	0.56
XGBoosted Forest	Syn. & Emo.	0.78	0.79	0.78	0.81	0.90	0.90	0.90	0.90
CNN (2,3,4)	GloVe	0.81	0.79	0.80	0.85	0.89	0.89	0.90	0.89
	GloVe & Syn. & Emo.	0.81	0.78	0.79	0.85	0.90	0.90	0.90	0.90
	GloVe & Syn. & Emo. & Prof.	0.81	0.77	0.78	0.84	0.91	0.91	0.91	0.91
CNN (2,3,4,5)	GloVe	0.83	0.78	0.80	0.85	0.89	0.89	0.89	0.89
	GloVe & Syn. & Emo.	0.80	0.77	0.78	0.84	0.89	0.89	0.89	0.89
	GloVe & Syn. & Emo. & Prof.	0.81	0.79	0.80	0.85	0.90	0.90	0.90	0.90

tative as the word embedding. The different results of *social* and *agency* imply that the prediction of *social* label might not rely on the text input, and it’s easier to predict. An additional experiment run to explore this was the addition of the top 1000 unigrams as features for the XGBoosted Forest, however this led to little (1-2%) to no increase in predictive power. To further explore the feature set, we incrementally add the syntactic features and emotional features, followed by the profile features to the CNN model. The results show that adding the syntactic and emotional features leads to a slight drop for *agency*. Though adding the profile features to the CNN model might lead to small improvements, we mainly focus on the word embedding, syntactic features and emotional features for the semi-supervised learning.

Semi-Supervised Learning. After getting the best models from the supervised learning, we generate the pseudo labels for the 72,324 unlabeled data using the XGBoosted Forest and the CNN models. Then we combine the labeled training data with the pseudo-label data to train the semi-supervised models via 10-fold cross-validation. The validation set is always held out during the training. Performance of our models are reported in Table 3.

Table 3: 10-fold cross-validation semi-supervised learning for *agency* and *social* prediction.

Model	Features	Agency					Social				
		P	R	F1	Acc	AUC	P	R	F1	Acc	AUC
XGBoosted Forest	Syn. & Emo.	0.80	0.81	0.79	0.81	0.68	0.91	0.91	0.91	0.91	0.91
CNN (2,3,4)	GloVe	0.81	0.78	0.79	0.85	0.78	0.89	0.89	0.89	0.89	0.89
	GloVe & Syn. & Emo.	0.80	0.79	0.79	0.84	0.79	0.90	0.90	0.90	0.90	0.90
CNN (2,3,4,5)	GloVe	0.82	0.79	0.80	0.85	0.79	0.89	0.89	0.89	0.89	0.89
	GloVe & Syn. & Emo.	0.80	0.78	0.79	0.84	0.78	0.90	0.90	0.90	0.90	0.90

We had expected performance improvements from semi-supervised learning, but notice that for the CNN model, the additional 70k pseudo-labeled data does not improve performance. Compare Table Table 2 and Table 3. In Table 2 for *agency* prediction, the

best model CNN region (2, 3, 4, 5) with GloVe has an F1-score of 0.80 and in Table 3 its F1-score remains 0.80. Similarly, the best CNN model with embeddings, syntactic and emotional features for *social* prediction gets an F1-score of 0.90 in Table 2 as well as Table 3.

Note also that the XGBoosted Forest provides good performance with syntactic and emotional features, and its performance improves slightly after semi-supervised learning, e.g. for *agency* prediction, it has an F1-score of 0.78 for supervised learning, and 0.79 for semi-supervised learning. For the *social* prediction, its F1-score is 0.90 for supervised learning and 0.91 for semi-supervised learning. These results encourage us to investigate the further impact of syntactic and emotional features for affective prediction tasks.

Shared Task Submission. We’ve submitted the supervised models of the XGBoosted Forest, CNN with GloVe, and CNN with syntactic and emotional features. As for the semi-supervised learning, we’ve submitted the XGBoosted Forest, CNN with GloVe, and CNN with syntactic and emotional features especially for region size (2, 3, 4).

3 Concepts Modeling

This work extends the modeling procedures described above to predicting the *concepts* label within the HappyDB. We are interested in the *concepts* features since they represent the theme of different types of happy moments. However, we expect that this is a much more difficult task as it is a multi-class problem. For the *concepts* modeling task, we are interested in both improving the prediction results, and interpreting the performance of the models.

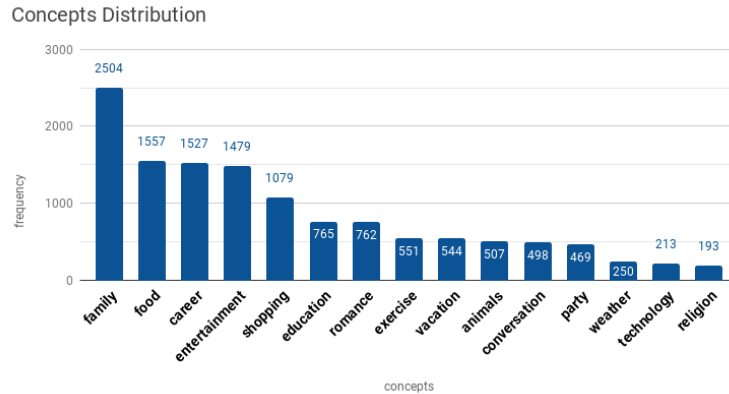


Fig. 2: The frequency of *concepts* is ordered from highest to lowest from left to right. For example, The most common *concepts* is Family with 2504 examples and the least common *concepts* is Religion with 193 examples.

3.1 XGBoosted Forest Model

For this task the labeled 10k data set is split into a training set containing 67% of the data and the remaining 33% is used for the test set. Within the training set, a 10-fold cross-validation procedure is used.

There are 15 unique *concepts* in the corpus, which are shown in Figure 2, however they are commonly associated with each other as some instances within HappyDB have multiple *concepts* attached. To simplify and examine if *concepts* are distinguishable from each other, we only model the cases where a singular *concepts* tag has been applied. Using the same feature set and modeling procedure for the XGBoosted Forest, Table 4a shows the performance of the model on each unique *concepts* tag. The rows of the table are ordered by model performance.

Overall the model shows some promising performance across all *concepts*. The model shows some good performance in the top 3 *concepts*, despite the small sample size for Religion, it appears to be performing the best. However all other *concepts* with lower than 100 instances show much poorer performance.

One possibility for the poor performance in some of these *concepts* may be the association between them that is already present within the HappyDB, as many *concepts* are used together. Future work could take these common associations and look at potentially making a more hierarchical modeling procedure.

3.2 CNN Model

Table 4: Concepts Prediction. Concepts are ordered by decreasing F1-score in each case.

(a): XGBoosted Forest

Concept	P	R	F1
Religion	0.86	0.95	0.90
Food	0.81	0.86	0.83
Family	0.73	0.90	0.81
Career	0.67	0.80	0.72
Entertainment	0.70	0.75	0.72
Shopping	0.67	0.66	0.66
Animals	0.59	0.58	0.58
Romance	0.65	0.48	0.56
Conversation	0.56	0.52	0.54
Weather	0.54	0.41	0.47
Education	0.54	0.40	0.46
Party	0.71	0.33	0.45
Exercise	0.48	0.36	0.42
Vacation	0.50	0.34	0.40
Technology	1.00	0.11	0.20

(b): CNN

Concept	P	R	F1
Religion	0.98	0.91	0.94
Food	0.91	0.86	0.88
Entertainment	0.88	0.86	0.87
Animals	0.92	0.84	0.87
Career	0.85	0.87	0.86
Shopping	0.88	0.84	0.86
Family	0.86	0.83	0.84
Education	0.88	0.73	0.79
Weather	0.88	0.72	0.77
Exercise	0.82	0.74	0.76
Party	0.89	0.70	0.76
Vacation	0.84	0.71	0.75
Conversation	0.82	0.71	0.75
Romance	0.77	0.64	0.68
Technology	0.83	0.59	0.64

Since the CNN model handles multiple classes, we convert the value of *concepts* into a one-hot vector with 15 dimensions, allowing multiple *concepts* attached to a happy moment. We explore the performance of CNN region size (2, 3, 4, 5) with syntactic features and emotional features by 10-fold cross-validation. The overall accuracy of the model is 0.596 and F1-score is 0.629. The metrics for each *concepts* are demonstrated in Table 4b.

Table 4b shows that the CNN model is generally better than the XGBoosted Forest Model for the *concepts* prediction. For example, the highest F1 for CNN is 0.94 for Religion, and the lowest F1-score is 0.64 for Technology; while the highest F1-score for XGBoosted Forest is 0.9 and lowest is 0.2, meaning that the CNN model is more robust and steady for multi-class prediction. The *concepts* that are improved more than 30% by CNN are Weather, Party, Education, Exercise, Vacation, and Technology. We suggest that the different performances are caused by the word information that is missing in XGBoosted Forest. The POS and LIWC features used within the XGBoosted Forest appear to be sufficient enough to cover general patterns only within Religion, Food, Entertainment, Career, and Family.

Besides the above features, we also explored adding the profile features. Its overall accuracy is 0.601 and F1-score is 0.626. Adding the profile features to the model doesn't provide large improvements, but there are small improvements (1%) for most of the *concepts*, with the biggest improvement in Exercise which increased by 3%. Intuitively, some profile features can be good markers for *concepts* prediction such as *age*, *married*, and *parenthood*. For instance, young people tend to discuss events of Education, while the parents are likely to be happy for a Family theme. On the other hand, the *concepts* that drop 1% after adding the profile features are Shopping, Weather, Party, Conversation. The biggest changes include Technology, whose F1-score drops by 4%.

Though these models have different performance, they all illustrate the difficulty of predicting certain *concepts* and share a similar prediction trend. The trend is not affected by the size of the training data as in Figure 2. For example, both models agree that Religion and Food are much easier to predict than Romance and Technology, but both Religion and Technology have small training sets, suggesting that perhaps Religion contains many discriminative patterns that make it easier to predict. The performance also implies that Religion and Food might have some distinctive patterns with global agreement to represent the happy moment while Romance and Technology might vary, meaning that people might have very different views of what causes happiness in these *concept* themes.

3.3 Syntactic Pattern Analysis

To further interpret the performance of the above models, we apply AutoSlog [8, 12], a weakly supervised linguistic-pattern learner, to collect the compositional syntactic patterns for the 10k labeled data. Table 6 illustrates the most frequent syntactic patterns in the data. For each pattern, we list the top 3 *concepts* with the highest probability (no less than 10%) given the pattern. In the top 15 list, there are 3 patterns include MY (FAMILY MEMBER). This might explain why *social* prediction is easier than the other tasks. As for the *concepts*, a Family theme usually dominates the pattern of MY

(FAMILY MEMBER), which implies that when Romance and Family co-occur, the classifier would tend to predict Family, leading to a low recall for Romance.

Table 6: Selected top 15 syntactic patterns using AutoSlog-TS Templates.

Freq	Pattern and Text Match	Concepts Probability and Examples
1395	<subj> ActVp (WENT)	Family 0.15, Shopping 0.14, Food 0.12 <i>Example: I went for a walk with my wife.</i>
1303	<subj> ActVp (GOT)	Career 0.25, Family 0.17, Food 0.10 <i>Example: I finally got a job interview.</i>
1153	<subj> ActVp (MADE)	Family 0.23, Food 0.23, Career 0.13 <i>Example: I made a delicious meal.</i>
605	Subj AuxVp <dobj> (HAVE I)	Food 0.32, Career 0.12, Family 0.12 <i>Example: I had excellent dinner.</i>
571	<subj> AuxVp AdjP (BE HAPPY)	Family 0.23, Career 0.13 <i>Example: I was happy to see some friends while they were on vacation.</i>
518	Adj Noun (MY HUSBAND)	Family 0.44, Romance 0.27, Food 0.13 <i>Example: My husband surprised me with my favorite treats.</i>
500	<subj> AuxVp AdjP (BE ABLE)	Career 0.16, Family 0.13, Shopping 0.12 <i>Example: I was able to get off of work early.</i>
495	Adj Noun (MY WIFE)	Family 0.40, Romance 0.28, Food 0.13 <i>Example: I got a kiss from my wife!</i>
476	ActVp <dobj> (BOUGHT)	Shopping 0.61 <i>Example: I bought a new laptop.</i>
458	Adj Noun (MY FAMILY)	Family 0.51, Food 0.13, Vocation 0.12 <i>Example: I went on vacation with my family</i>

Moreover, the *concepts* of Family, Career, Food, and Shopping contain distinctive syntactic patterns, consistent with classification performance. To validate our assumptions about the discriminative level of Technology and Religion, we look at the most frequent patterns. For example, the most common pattern for Technology is (BOUGHT) with 36 examples, and Religion’s common pattern is (WENT TO) with 148 examples. Similar to lexical-diversity, we define the pattern-diversity as the number of unique patterns divided by the total number of patterns. Technology contains 1078 syntactic patterns, the average frequency for each pattern is 2, the standard deviation of frequency is 3, and the pattern-diversity is 0.55, whereas Religion includes 304 patterns, the average count is 3, the standard deviation is 10, and the pattern-diversity is 0.51. As for Family, which has general syntactic patterns, the average count is 3, the standard deviation is 10, and the pattern-diversity is 0.37. A larger standard deviation implies more typical patterns, and a smaller pattern-diversity implies the *concepts* tends to include stronger syntactic patterns. Since Technology has a similar number of examples as Religion,

this suggests that Religion is more easily identified because it has many typical syntactic patterns. Our observation suggests that the syntactic patterns can be strong markers for affective classification tasks even when the *profile features* are missing.

4 Discussion and Future Work

We explored the features and models with supervised learning and semi-supervised learning for *social*, *agency* and *concepts* in order to answer some of the questions during the experiments:

- *Whether syntactic features and emotional features, which are generated from the text, are representative?* Our experiments show that the syntactic features and emotional features are informative features as they are competitive to word embeddings, and could outperform the word embeddings for some tasks.
- *Whether the profile features are representative? How to convert them into a meaningful granularity?* The profile features include large background information of the writer, therefore, they might represent some general happiness group. However, some features should be mapped to a meaningful group, such as country and age. Though our experiment results are relatively low for the profile features, it could slightly improve the predictions of *social* and some *concepts* when combined with other features. On the other hand, the text input and its extracted features perform well on the prediction task, which indicates that the demographic information is not necessary to achieve decent performance.
- *Whether the neural network model outperforms the traditional machine learning model? Whether the best models for agency and social prediction give good performance for concepts prediction?* The CNN models provide promising performance for multi-class classification and semi-supervised learning, while the traditional machine learning method XGBoosted Forest also generates competitive or even better results for binary class prediction and semi-supervised learning. Our best feature sets and models for *social* and *agency* prediction also provide good performance for *concepts* prediction. Our syntactic pattern analysis also demonstrates that these tasks share common characteristics. Therefore, we believe that they are generally robust models for affective classification tasks.

During the experiments, we realize that the variation and the definition of happy moments might affect the performance of *concepts* modeling, and imply the generalization level of different *concepts*. Some generic characteristics, which are implied by the syntactic and emotional features, are shared between the classes of *agency*, *social* and *concepts*. The linguistic-pattern learner AutoSlog provides insightful syntactic patterns for us to interpret the models. In future work, we will focus on utilizing syntactic patterns for other affective classification tasks and identify common patterns. We also hope to explain such patterns and the *concepts* theme with psychological theories. Finally, we are curious to see if there are generic characteristics or common compositional semantic patterns for modeling happy or unhappy moments with cross-domain data.

References

1. Asai A.& Evensen S.& Golshan B.& Halevy A.& Li V., Lopatenko A.& Stepanov D.& Suhara Y.& Tan W.C.& Xu Y.: Happydb: A corpus of 100,000 crowdsourced happy moments. In Proceedings of LREC 2018. European Language Resources Association(ELRA) (2018)
2. Chen, T. & Guestrin, C.: Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. ACM (2016)
3. Compton, R.& Chen, J.& Haber, E.& Badenes, H.& Whittaker, S.: Just the Facts: Exploring the Relationship between Emotional Language and Member Satisfaction in Enterprise Online Communities. In Proceedings of the 11th International Conference on Web and Social Media. AAAI. (2017) <https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15664>
4. Jaidka, K.& Mumick,S. & Chhaya, N. & Ungar, L., : The CL-Aff Happiness Shared Task: Results and Key Insights. In Proceedings of the 2nd Workshop on Affective Content Analysis @ AAAI (AffCon2019). Honolulu, Hawaii (2019)
5. Mohammad, S. M. & Turney, P. D.: Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, pp. 26-34. Association for Computational (2010) Linguistics.
6. Pennington, J., Socher, R. & Manning, C. D.: GloVe: Global Vectors for Word Representation. In Proceedings of EMNLP 2014, pp. 1532-1543. (2014)
7. Reed, L. & Wu, J. & Oraby, S. & Anand, P. & Walker, M.: Learning Lexico-Functional Patterns for First-Person Affect. Association for Computational Linguistics (ACL) (2017)
8. Riloff, E.: Automatically generating extraction patterns from untagged text. pp. 1044-1-49. Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96) (1996)
9. Tausczik, Y. R. & Pennebaker, J. W.: The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), pp 24-54. (2010)
10. Toutanova, K.& Klein, D.& Manning, C. D., & Singer, Y. (2003, May). Feature-rich part-of-speech tagging with a cyclic dependency network. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, pp. 173-180. Association for Computational Linguistics. (2015)
11. Wilson, T. & Hoffmann, P. & Somasundaran, S. & Kessler, J.& Wiebe, J.& Choi, Y. & Patwardhan, S.: OpinionFinder: A system for subjectivity analysis. In Proceedings of hlt/emnlp on interactive demonstrations. pp. 34-35. Association for Computational Linguistics (2005)
12. Wu, J. & Walker, M. & Anand, P. & Whittaker, S.: Linguistic Reflexes of Well-Being and Happiness in Echo. The 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA) (2017), Empirical Methods in Natural Language Processing (EMNLP) (2017)
13. Zhang, Y. & Wallace, B. C.: A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. IJCNLP (2015)