

# Time Emotional Analysis of Arabic Tweets at Multiple Levels

Amr M. Sayed, Samir AbdelRahman, Reem Bahgat and Aly Fahmy  
Computer Science Department  
Faculty of Computers and Information  
Cairo University  
Giza, Egypt

**Abstract**—Sentiment and emotional analyses have recently become effective tools to discover peoples attitudes towards real-life events. While Many corners of the emotional analysis research have been conducted, time emotional analysis at expression and aspect levels is yet to be intensively explored. This paper aims to analyse people emotions from tweets extracted during the Arab Spring and the recent Egyptian Revolution. Analysis is done on tweet, expression and aspect levels. In this research, we only consider surprise, happiness, sadness, and anger emotions in addition to sarcasm expression. We propose a time emotional analysis framework that consists of four components namely annotating tweets, classifying at tweet/expression levels, clustering on some aspects, and analysing the distributions of people emotions, expressions, and aspects over specific time. Our contribution is two-fold. First, our framework effectively analyzes people emotional trends over time, at different fine-granularity levels (tweets, expressions, and aspects) while being easily adaptable to other languages. Second, we developed a lightweight clustering algorithm that utilizes the short length of tweets. On this problem, the developed clustering algorithm achieved higher results compared to state-of-the-art clustering algorithms. Our approach achieved 70.1% F-measure in classification, compared to 85.4% which is the state of the art results on English. Our approach also achieved 61.45% purity in clustering.

**Keywords**—Emotional Analysis; Sentiment Analysis; Clustering; Two-Step Classification; Time Analysis

## I. INTRODUCTION

Nowadays, social media become interactive environments that present people opinions, emotions, and thoughts regarding specific daily events and aspects. Recently, sentiment and emotional analyses have become very effective mining tools to extract potential and accurate on-time information from tweets [1], [2], [3], [4]. Well-known social media researches were developed to classify, extract, or retrieve information at documents, sentences, expressions, aspects, or words according to some sentiments or emotions [5].

In this research, we are interested in the first three years of Arab Spring - Egyptian Revolution and the related human feelings (emotions) namely: surprise, happiness, sadness, anger, and sarcasm emotion. As well, we identified seven main aspects in that period as explained in the Data Collection and Annotation Section. As our main contribution, we propose a framework to analyze people emotional trends through the aforementioned period at different fine-granularity levels of tweets, expressions, and aspects. The framework comprises four main components namely annotating tweets, classifying

at tweet/expression levels, clustering on some aspects, and analyzing the distributions of people emotions, expressions, and aspects over that time period.

In the classification component, we utilized Conditional Random Fields (CRF) and AdaBoostMH classifiers for classifying emotions at tweets and expression levels in which CRF reported better results. In the clustering component, we evaluated K-Nearest Neighbor (K-NN), Expectation Maximization (EM), and Latent Dirichlet Allocation (LDA) to cluster the tweets with respect to the aspects included in this paper. Moreover, we proposed Tweets Lightweight Clustering (TLC) algorithm that utilizes the tweet short-length nature to achieve better performance. TLC resulted in the highest purity and the lowest Kullback–Leibler Divergence in comparison to the mentioned clustering algorithms. In the time and aspect analysis component, we used Bayesian rules to calculate the probability and cumulative distributions of the emotions, aspects, and expressions over time. As well, we used Pointwise Mutual Information (PMI) to measure the dependency between two emotional expressions.

The rest of this paper is organized as follows. In the second section, some main English and Arabic related works are highlighted. The third section covers our framework details. The specifications and analysis of our experiments are described in the fourth section. We finally pinpoint the paper with main conclusion and future work.

## II. RELATED WORK

Despite a lot of research have been conducted in English Language, research in Arabic language is considered immature and developing. Further, less attention has been devoted to the emotional analysis area specially on Arabic language. In the following, we summarize the relevant sentiment and emotional analysis research:

**Sentiment Analysis:** [6] used n-gram to learn SVM classifier at tweet level to classify Egyptian dialect tweets. [7] used a naive Bayesian classifier on Arabic dialect tweets having the best F-measure for positive class. [8] used logistic regression model to build a sentiment classifier to classify English tweets on expression-level. [9] used deep convolutional neural network to classify the polarity of English sentences. [10] used SVM classifier with semantic and syntactic features learnt by multi-dialectal Arabic tweets.[11] used SVM classifier with Arabic lexicons and n-gram for sentiment classification at tweet level as well as aspect level. [12] classified English

brand-related tweets using Dynamic Architecture for Artificial Neural Networks.

**Emotional Analysis:** [13] used syntactic and lexical features and a Vector Space Model (VSM) to classify limited types of documents into six classes and got the best F-score on the happiness class. also [14] used Naive Bayes, SVM, Hyperpipes and Voting Feature Intervals to classify Arabic poems into four classes. [15] used syntactic feature to train SMO and SVM classifiers to classify Arabic tweets into six classes and got the best F-score on the Fear class. [3] used hash-tags and logistic regression to classify English tweets into five emotions having the best F-measure on the affection class.[16] used SVM and VSM to classify English documents into seven emotions and got the best F-score on fear class.

### III. THE PROPOSED TIME EMOTIONAL ANALYSIS FRAMEWORK

In the following, we present the framework classification and clustering components as three paragraphs discussing the extracted features, techniques, and the related evaluation criteria. The time and aspect analysis component is presented as a set of equations used to study distributions of emotions and expressions overtime per aspect. In the rest of this paper, Arabic text is quoted between double quotes of “ ” and its English translation is presented between rectangular brackets [ ].

#### A. Data Collection and Annotation

We used Ekman’s set of emotions [17] in classification but with some modifications. We considered anger and disgust as one class and the same for sadness and fear. We have also added the sarcasm class to our set of classes due to the high rate of sarcastic expressions in the collected data. We constructed a lexicon of 563 emotional Arabic words from twitter. We manually classified them into emotional words of happiness (41), surprise (173), sadness (66) and anger (44), sarcasm (16), neutral (103), and multi-emotional (120). Examples of the collected words are: “مبسوط” [glad] happiness, “بيقولك” [someone said] sarcasm, “استغرب” [Surprised] surprise, “يرحمه” [Bless his soul] sadness, “فأشل” [Unstuck] anger, “نفسى” [I hope] multi-emotional and “حائس” [I feel] neutral. We collected Arabic tweets written in Egyptian dialect over the period January 1 2011 through December 31 2013. Each collected tweet should have at least one of seven keywords (aspects) presenting influential figures or events in this period. The keywords are namely: “ثورة ٢٥ يناير” [25 Jan Revolution], “ثورة ٣٠ يونيو” [30 June Revolution], “برنامج البرنائج” [El-Bernameg TV Show], “عبد الفتاح السيسي” [Abdel Fatah El-Sisi], “حسني مبارك” [Hosni Mubarak], “محمد مرسي” [Mohamed Morsi] and “الأخوان المسلمين” [Muslim Brotherhood]. Tweets were pulled from twitter search engine. Each tweet contained the user-name, the date, and the associated text. We removed any tweet that contained only URLs and the related replicates (if any). As a result of the previous steps, a corpus of 111,413 tweets was built. Based on the mentioned seven keywords or aspects, we had the related clusters of 6886, 2949, 5736, 11698, 17745, 41075, and 44604 tweets respectively. Of note, we may find a tweet replicated in two or more clusters if it has two or more keywords in common. To build the classification

model, 10,177 tweets were selected randomly from that corpus and all words were normalized. Normalization achieved by removing all non-Arabic characters from the text. These tweets were annotated manually by three specialized linguists and also revised by the authors of the paper. It was made of 608, 1361, 3368, 1056, 1688, 207, 1008 tweets from happiness, sarcasm, surprise, sadness, anger, neutral, and multi-emotional classes respectively. To avoid confusion that might occur during the annotation process among annotators, we used the following criteria: (1) The annotator should not involve his personal feelings towards the matter in concern during the annotation process. (2) To consider a tweet for the annotation, it should contain at least one of the following three items: emotional expression, emotional term (LOL, HHHHH, etc...), and/or emotional symbols ( :), :(, etc...). (3) Any tweet that contains non-Egyptian dialect was discarded. This annotated corpus was used to develop the classification model which was used to automatically annotate the rest of the 111,413 tweets.

The processes of tweets collection and annotations lasted for around six months. Of note, we aim at releasing the data publicly. As well, since our classifier is effective and it was used to annotate unseen data for clustering component in a systematic manner, it will be very effective and doable to annotate any other amount of Egyptian tweets having similar emotions and aspects. Furthermore, the annotation process could be easily extended to include other emotions and aspects.

#### B. Classification Component

This component aims at enabling the framework to classify the tweet emotions at levels of expressions and tweets.

**Classifiers:** In the classification component, we used the two-step classification approach adapted from [18]. At the expression level, we built the one-step and the two-step classifiers using Conditional Random Fields and AdaBoostMH baselines for each step. To classify emotions at tweet level, the following criteria were applied: (1) If a tweet has no clue or it has neutral clues only, then it’s considered a neutral tweet. (2) If a tweet contains more than one emotional clue from a specific class, it’s considered an emotional tweet that holds emotion from that class. (3) If a tweet contains more than one emotional clue from different classes, then it’s considered a multi-emotional tweet.

**Classification Features:** Following [19], we used four types of features groups with adaptation to Arabic language specially Egyptian dialect. For example we used the Egyptian negation words “مش” [not] and “مكانش” [wasn’t]. These features are: (1) word features such as the word itself, part-of-speech, prior polarity, strength and a negation checker, (2) modification features that are related to the context in which the word appears to indicate if a word is preceded by adjective, adverb or intensifier and also to indicate if a word modifies or is modified by a subjective clue, (3) tweet features are counters for strong and weak clues in the context in addition to morphological counters, and (4) structure features that consider the tweet structure and the relations among its words extracted from the Stanford parser [20] dependency parse tree.

**Evaluation Criteria:** In our experiments, 10-fold cross validation criterion was applied. We used three performance measures for the classification evaluation: (1) **precision**

that is the percentage of the retrieved emotional clues to the relevant ones, (2) **recall** that is the percentage of relevant emotional clues to retrieved ones, and (3) **F-measure** that is a measure of both precision and recall (equation 1).

$$F - measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (1)$$

### C. Clustering Component

This component aims at enabling the framework to partition tens thousands of tweets according to their aspects' similarities. This step is necessary to analyze the aspects' features in the subsequent component.

**Clustering Techniques:** We evaluated three different clustering algorithms namely *K-Nearest Neighbors (K-NN)* as a baseline, *Expectation Maximization (EM)*, and *Latent Dirichlet Allocation (LDA)*. Our evaluation results showed that: (1) They acquired in general fair results for highly overlapping clusters. (2) The resultant clusters' distributions significantly deviated from the gold standard ones. This clustering performance motivated us to use a sort of bi-gram topic model and we named it *Lightweight Clustering (TLC)* in which each cluster is identified by the most frequent uni/bi-gram and the related subsequent 100 co-occurring uni/bi-grams as bag-of-words. The algorithm could be summarized as follows. In step 1, we generate uni-gram and bi-gram with their frequencies (we excluded all possible stop words). In step 2, we select top frequent (m) uni/bi-grams to present (m) clusters such that: (1) each gram presents one cluster, (2) all clusters have no grams in common (e.g., "عبد الفتاح السيسي" and "السيسي" should not be assigned to different clusters), and (3) they co-occur together in  $> 70\%$  of the tweets. In step 3, we assign each of the subsequent 100 grams to one of the (m) clusters as related bag-of-words such that the candidate gram co-occur with the cluster gram in  $> 70\%$  of the tweets. In step 4, we assign each tweet to a cluster where at least one of the tweet's grams appears in the corresponding cluster bag-of-words grams. Of note, a tweet could be assigned to more than one cluster. In all algorithms, we used  $m = 7$  presenting the 7 aspects. In addition, the parameters of our algorithm were selected as the best effective thresholds based on extensive experimentations on this corpus.

**Clustering Features:** We conducted extensive feature evaluation for each clustering algorithm and the following are the best performing features. In K-NN algorithm, we used only the raw text of a tweet as the bag-of-words feature. For EM algorithm, a feature vector of the binary values was used to present the presence/absence of the most frequent uni-gram, bi-gram, and tri-gram in the tweet. The LDA algorithm uses only one type of features which is a pair of word and its frequency in the given tweet. To overcome this limitation, we combined each two/three words with a delimiter ("-") to present the related bi/tri-gram level in addition to the uni-gram. We used only the most frequent 20 grams as features for EM and LDA algorithms.

**Evaluation Criteria:** To evaluate the quality of each clustering algorithm, we used the cluster purity that is the

percentage of the most frequent class in that cluster. **Kullback-Leibler Divergence ( $D_{KL}$ )** was used to compare how close the algorithm generated word distribution in the clusters is to that of the gold standard; the lower the KL-divergence is, the closer to that of gold standard one is. KL-divergence uses the following equation to quantify the difference between two probability distributions A and T, where T presents the true distribution (gold standard) of the clustering data and A presents the algorithm approximation of T.

$$D_{KL}(T||A) = \sum_{i=1}^n T_i \cdot \log\left(\frac{T_i}{A_i}\right) \quad (2)$$

### D. Time and Aspect Analysis

This component is important to show a meaningful analysis on tweets using the classification and clustering components. Analysis on tweets was applied at several levels to generate the most useful and important information. To achieve the component goal, the following equations are used:

- Cumulative probability of emotion (E) given a time interval (T) (with start (s) and end (t)) and a specific cluster (C):

$$P(E|C, T) = \sum_{i=s}^t P(E|C, i) \quad (3)$$

- Probability of emotion (E) given set of tweets (S) of specific time (i):

$$P(E|S_i) = \frac{\# \text{ tweets with emotion E in S}}{\# \text{ tweets in S}} \quad (4)$$

- Probability of emotional expression (Ex) given tweets of a specific cluster (C):

$$P(Ex|C) = \frac{\# \text{ tweets with expression Ex in C}}{\# \text{ tweets in C}} \quad (5)$$

- Pointwise mutual information (PMI) which measures the dependency between two emotional expressions  $Ex_i$  and  $Ex_j$ :

$$PMI(Ex_i, Ex_j) = \frac{P(Ex_i|Ex_j)}{P(Ex_i)} \quad (6)$$

## IV. EXPERIMENT RESULTS AND ANALYSIS

### A. Classification Experiments

Our experiments were confirmed [18] conclusion that the two-step classifier results were better than those of the one-step one. Table 1-4 list the results of our two-step classifiers. We can see that multi-emotional class has the lowest scores because multi-emotional tweets are rare and not rich in expressions since the maximum length of a tweet is only 140 letters. Sarcasm class achieved the highest scores due to the significant availability of tweets and the clarity of expressions. The presence of some expressions that can be used the same way in sad and angry contexts led to their weak scores. Neutral class achieved poor results due to the use of emotional expressions and emotion symbols by writers in non-emotional contexts.

TABLE I: Two-Step Step(A) Expression-Level Classifiers Results

Model	CRF Classifier			AdaBoostMH Classifier		
	precision	recall	F-measure	precision	recall	F-measure
Emotional	91.3%	78.3%	84.3%	89.4%	99.0%	94.0%
Neutral	88.1%	46.8%	60.9%	49.5%	07.4%	12.9%
<b>Average</b>	<b>89.70%</b>	<b>62.55%</b>	<b>72.60%</b>	<b>69.45%</b>	<b>53.20%</b>	<b>53.45%</b>

TABLE II: Two-Step Step(B) Expression-Level Classifiers Results

Model	CRF Classifier			AdaBoostMH Classifier		
	precision	recall	F-measure	precision	recall	F-measure
happiness	50.8%	52.1%	51.0%	43.0%	35.9%	38.5%
Sarcasm	63.5%	<b>74.9%</b>	<b>68.1%</b>	58.9%	<b>71.6%</b>	<b>64.1%</b>
Surprise	60.7%	57.4%	58.7%	58.0%	59.2%	58.3%
Sadness	<b>64.4%</b>	66.3%	64.6%	<b>61.4%</b>	59.2%	60.0%
Anger	61.2%	52.5%	56.1%	60.3%	49.6%	54.1%
Multi-emotional	00.2%	05.8%	01.0%	00.0%	00.0%	00.0%
<b>Average</b>	<b>50.13%</b>	<b>51.50%</b>	<b>49.92%</b>	<b>46.93%</b>	<b>45.92%</b>	<b>45.83%</b>

TABLE III: Two-Step Step(A) Tweet-Level CRF Classifier Results

Measure	Precision	Recall	F-measure
Emotional	91.20%	84.20%	87.50%
Neutral	27.30%	42.20%	32.50%
<b>Average</b>	<b>59.25%</b>	<b>63.20%</b>	<b>60.00%</b>

TABLE IV: Two-Step Step(B) Tweet-Level CRF Classifier Results

Measure	Precision	Recall	F-measure
happiness	58.1%	43.6%	49.5%
Sarcasm	70.0%	<b>71.6%</b>	<b>70.1%</b>
Surprise	68.4%	53.8%	60.0%
Sadness	<b>70.1%</b>	63.5%	66.0%
Anger	66.1%	45.4%	53.4%
Multi-emotional	5.7%	31.5%	9.3%
<b>Average</b>	<b>56.40%</b>	<b>51.57%</b>	<b>51.38%</b>

### B. Clustering Experiments

Table 5 shows the purities of the clustering algorithms in which the highest values are in bold. Our algorithm, TLC, acquired the best purity on average with only one cluster (30 June Revolution) having the poorest purity. KL-divergence results are 0.043, 0.249, 0.252 and 0.286 for TLC, K-NN, LDA and EM clustering algorithms respectively. TLC showed the nearest distribution of all aspects to the gold standard ones. One may have the following notes. First, LDA acquired the worst purity results since the whole aspects were scattered through the seven clusters and hence there was no clear aspect per each cluster. Second, 30 June Revolution cluster showed the worst cluster purity results because it contained most of other aspects from the other clusters; it is a highly overlapped cluster. Third, although EM obtained higher purity than that of K-NN, it reported the worse KL-divergence since the distribution of all aspects per estimated cluster were not near to those of gold standard one. Finally, TLC resulted in balanced distributions among all main aspects (the most frequent class in the cluster) and that was showed as a good average purity (~61%) and the lowest KL-divergence (0.043).

### C. Time and Aspect Analysis

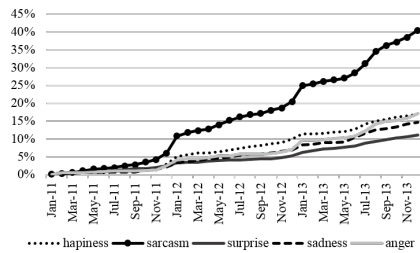
Analysis model was applied on all aspects. In Figures 1-7, we show three different views for each aspect: (a) the cumulative probability of each emotion on a specific time interval, (b) the probability of each emotion given each month tweets, and (c) the most frequent used emotional expressions. We can see many peaks in each emotion line presenting significant events that affect people feelings towards the aspect. As examples, we demonstrate the feelings (emotions) among Egyptians through the revolutions of 25 January and 30 June overtime as follows.

The happiness emotion increased with the beginning of the revolution in January 2011 due to the happiness of the Egyptian people of the successful completion of the revolution. It was also a high curve throughout the year and a half until mid-2012 due to the moving-forward steps of the revolution towards its goals, e.g. the handover of power, the parliamentary elections and the presidency, and the handover of power to the first elected president. The surprise emotion seemed clear in February and March due to the unexpected action of Mubarak when he delivered the power and that the revolution succeeded in its first step. Surprise was also featured in the month of July 2011 because some people attributed the revolution for themselves. The sadness emotion was evident in October 2011 due to the slowdown in achieving the goals of the revolution. The anger emotion appeared clearly in February 2011 because of the violence and the killings of protesters during the revolt against the regime.

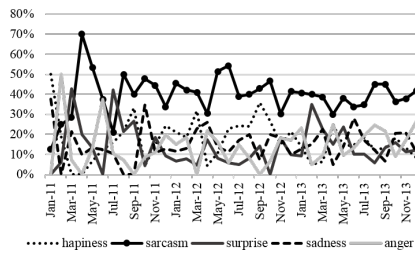
A high bending of the happiness emotion began in June 2013 due to the rapid success of the revolution and then began to decline, and increased again in September 2013 because of the happiness of the Egyptians for their ability to maintain the revolution and the beginning of the moving-forward steps towards the revolution goals. The levels of the surprise emotion were not high due to the absence of unexpected events. The grief emotion showed high levels in August 2013 because of the coming together of the people and the emergence of the post-revolution demanding the return of the former regime groups. Anger surfaced at the end of 2013 in sync with the joy due to the conflict situation in Egypt.

TABLE V: 7-Cluster Clustering Results

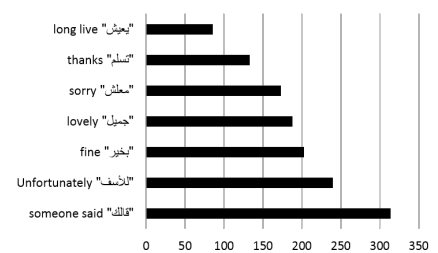
Cluster	KNN	EM	LDA uni-gram	LDA bi-gram	LDA tri-gram	TLC
25 Jan Revolution	01.59%	14.62%	01.03%	06.96%	04.63%	<b>71.98%</b>
30 Jun Revolution	00.43%	00.08%	02.13%	<b>12.04%</b>	03.75%	06.71%
El-Bernameg TV Show	08.87%	06.03%	18.82%	01.12%	08.32%	<b>68.98%</b>
Abdel Fatah El-Sisi	13.60%	<b>95.70%</b>	18.78%	07.49%	07.62%	66.15%
Hosni Moubarak	11.24%	<b>99.87%</b>	31.66%	20.45%	12.49%	69.62%
Mohamed Morsi	84.33%	<b>84.79%</b>	53.06%	40.68%	31.76%	69.59%
Muslim Brotherhood	36.23%	<b>83.33%</b>	63.41%	35.92%	35.18%	77.10%
<b>Average Purity</b>	<b>22.47%</b>	<b>54.83%</b>	<b>26.98%</b>	<b>17.81%</b>	<b>14.82%</b>	<b>61.45%</b>



(a) Cumulative Probability

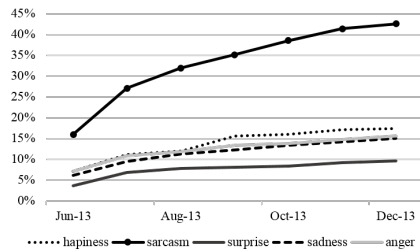


(b) Monthly probability

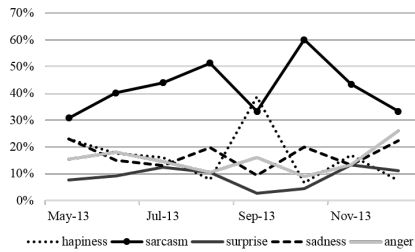


(c) Frequent Expressions

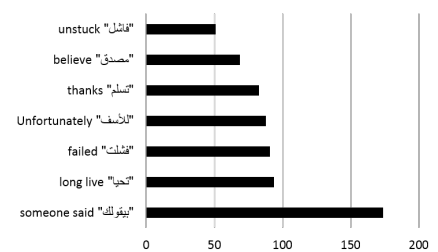
Fig. 1. 25 Jan Revolution



(a) Cumulative Probability

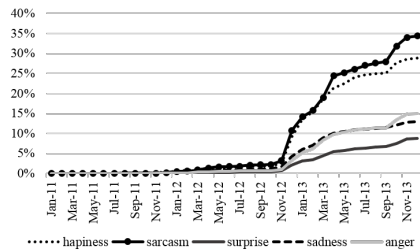


(b) Monthly probability

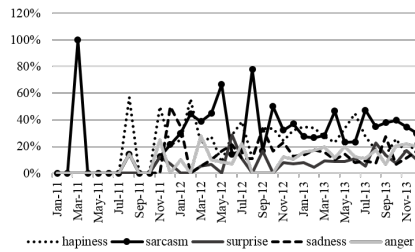


(c) Frequent Expressions

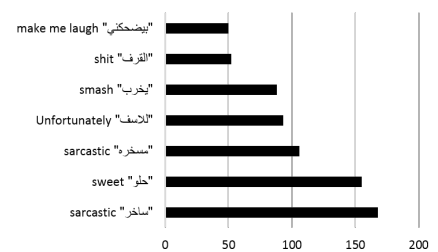
Fig. 2. 30 Jun Revolution



(a) Cumulative Probability



(b) Monthly probability



(c) Frequent Expressions

Fig. 3. El-Bernameg TV Show

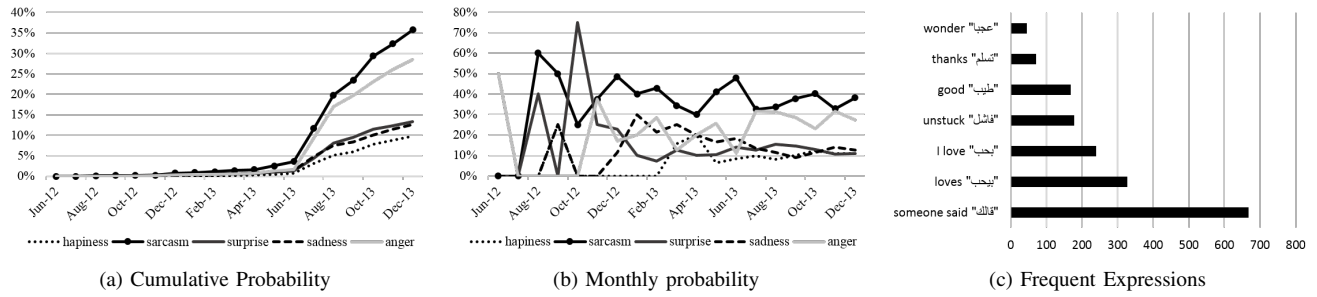


Fig. 4. Abdel Fatah EL-Sisi

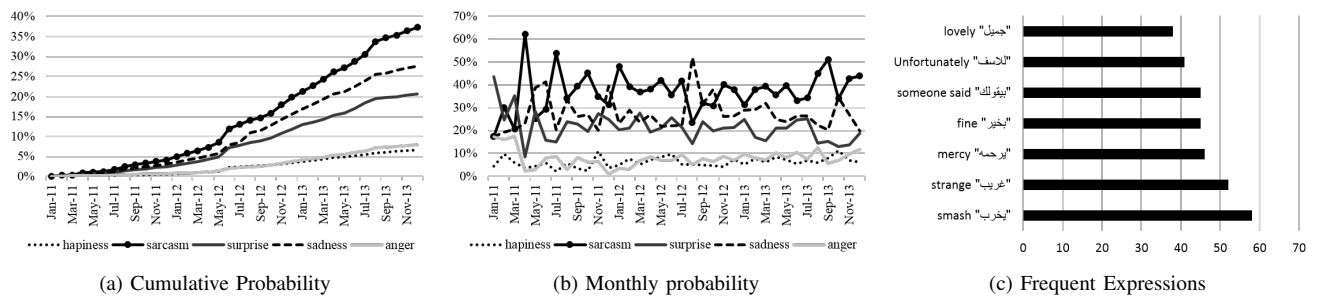


Fig. 5. Hosni Mubarak

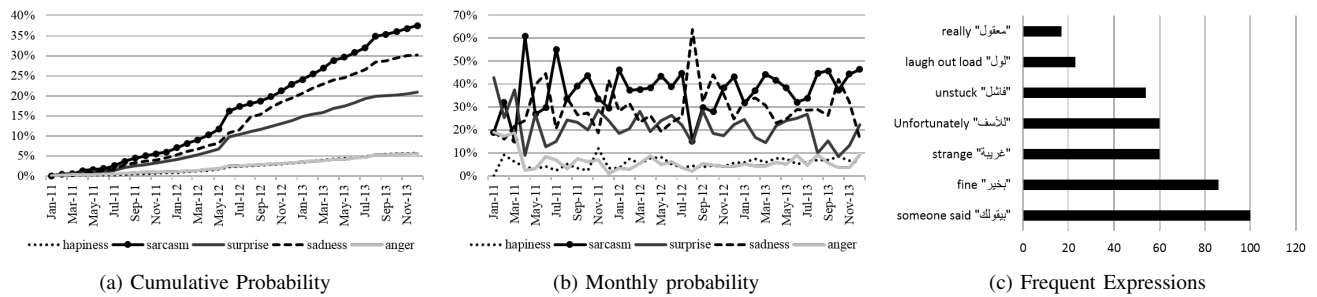


Fig. 6. Mohamed Morsi

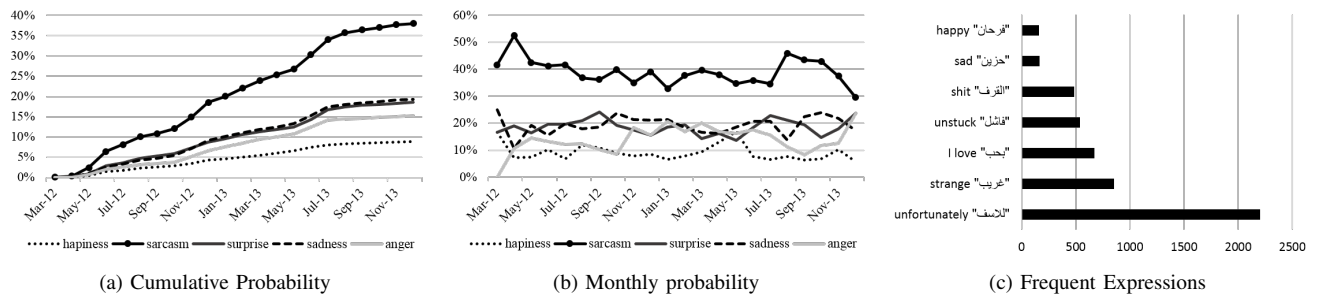


Fig. 7. Muslim Brotherhood

TABLE VI: PMI Measure Examples

Expression(1)	Expression(2)	PMI
“تخامة” [His Excellency]	“تحية طيبة” [Greetings]	10.61
“مقرف” [Nasty]	“مقرف” [Disgusting]	10.36
“يرحمه” [Bless his soul]	“الصبر” [Patience]	8.38
“غضبان” [Angry]	“غبي” [Idiot]	5.65
“استغرب” [Surprised]	“خسارة” [Loss]	2.28
“فأشل” [Unstuck]	“شكرا” [Thanks]	-2.67

PMI measure shows the degree of interdependence between emotional expressions and each others, as a high score means high bonding strength and vice versa as shown in table 6.

## V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a framework to effectively analyze people emotional trends over time at different fine granularity levels (tweets, expressions, and aspects). It can be easily adapted to other languages due to the utilization of language independent features. We also developed a lightweight clustering algorithm that benefit from the short length of tweets, for better clustering effectiveness against the state-of-the-art algorithms. Since the algorithm depends on the language grams, it can be utilized in many languages other than Arabic. In the future, the annotated corpus will be freely released to the public as an Arabic natural language community resource. We are planning to expand the set of emotions targeted by the classification model and to apply our work on other domains and languages like English. Moreover, we will investigate how to potentially boost multi-emotional classification performance.

## REFERENCES

- [1] E. Kouloumpis, T. Wilson, and J. D. Moore, “Twitter sentiment analysis: The good the bad and the omg!” *Icwsn*, vol. 11, pp. 538–541, 2011.
- [2] S. Mohammad, X. Zhu, and J. Martin, “Semantic role labeling of emotions in tweets,” *Proc. of WASSA*, pp. 32–41, 2014.
- [3] A. Qadir and E. Riloff, “Learning emotion indicators from tweets: Hashtags, hashtag patterns, and phrases.” in *EMNLP*, 2014, pp. 1203–1209.
- [4] H. Saif, Y. He, M. Fernandez, and H. Alani, “Contextual semantics for sentiment analysis of twitter,” *Information Processing & Management*, vol. 52, no. 1, pp. 5–19, 2016.
- [5] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press, 2015.
- [6] A. Shoukry and A. Rafea, “Preprocessing egyptian dialect tweets for sentiment mining,” in *The Fourth Workshop on Computational Approaches to Arabic Script-based Languages*, 2012, p. 47.
- [7] A. Mourad and K. Darwish, “Subjectivity and sentiment analysis of modern standard arabic and arabic microblogs,” in *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, 2013, pp. 55–64.
- [8] S. Rosenthal and K. McKeown, “Columbia nlp: Sentiment detection of subjective phrases in social media,” in *Second Joint Conference on Lexical and Computational Semantics (\*SEM)*, vol. 2. Citeseer, 2013, pp. 478–482.
- [9] C. N. dos Santos and M. Gatti, “Deep convolutional neural networks for sentiment analysis of short texts.” in *COLING*, 2014, pp. 69–78.
- [10] E. Refaee and V. Rieser, “Can we read emotions from a smiley face? emoticon-based distant supervision for subjectivity and sentiment analysis of arabic twitter feeds,” in *5th International Workshop on Emotion, Social Signals, Sentiment and Linked Open Data, LREC*, 2014.
- [11] J. Borge-Holthoefer, W. Magdy, K. Darwish, and I. Weber, “Content and network dynamics behind egyptian political polarization on twitter,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 2015, pp. 700–711.
- [12] D. Zimbra, M. Ghiassi, and S. Lee, “Brand-related twitter sentiment analysis using feature engineering and the dynamic architecture for artificial neural networks,” in *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2016, pp. 1930–1938.
- [13] A. F. El Gohary, T. I. Sultan, M. A. Hana, and M. El Dosoky, “A computational approach for analyzing and detecting emotions in arabic text,” *International Journal of Engineering Research and Applications (IJERA)*, vol. 3, pp. 100–107, 2013.
- [14] O. Alsharif, D. Alshamaa, and N. Ghneim, “Emotion classification in arabic poetry using machine learning,” *International Journal of Computer Applications*, vol. 65, no. 16, 2013.
- [15] O. Rabie and C. Sturm, “Feel the heat: Emotion detection in arabic social media content,” in *The International Conference on Data Mining, Internet Computing, and Big Data (BigData2014)*. The Society of Digital Information and Wireless Communication, 2014, pp. 37–49.
- [16] N. Riahi and P. Safari, “Implicit emotion detection from text with information fusion,” *Journal of Advances in Computer Research*, vol. 7, no. 2, pp. 85–99, 2016.
- [17] P. Ekman, “Facial expression and emotion.” *American psychologist*, vol. 48, no. 4, p. 384, 1993.
- [18] T. Wilson, J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis,” *Computational linguistics*, vol. 35, no. 3, pp. 399–433, 2009.
- [19] J. Wiebe and E. Riloff, “Creating subjective and objective sentence classifiers from unannotated texts,” in *Computational Linguistics and Intelligent Text Processing*. Springer, 2005, pp. 486–497.
- [20] R. Socher, J. Bauer, C. D. Manning, and A. Y. Ng, “Parsing with compositional vector grammars.” in *ACL (1)*, 2013, pp. 455–465.