



# Emotional Concept Extraction Through Ontology-Enhanced Classification

Danilo Cavaliere<sup>(✉)</sup>  and Sabrina Senatore 

Dipartimento di Ingegneria dell'Informazione ed Elettrica e Matematica Applicata - DIEM, Università degli Studi di Salerno, Fisciano, SA, Italy  
{dcavaliere,ssenatore}@unisa.it

**Abstract.** Capturing emotions affecting human behavior in social media bears strategic importance in many decision-making fields, such as business and public policy, health care, and financial services, or just social events. This paper introduces an emotion-based classification model to analyze the human behavior in reaction to some event described by a tweet trend. From tweets analysis, the model extracts terms expressing emotions, and then, it builds a topological space of emotion-based concepts. These concepts enable the training of the multi-class SVM classifier to identify emotions expressed in the tweets. Classifier results are “softly” interpreted as a blending of several emotional nuances which thoroughly depicts people’s feeling. An ontology model captures the emotional concepts returned by classification, with respect to the tweet trends. The associated knowledge base provides human behavior analysis, in response to an event, by a tweet trend, by SPARQL queries.

**Keywords:** Sentiment analysis · Simplicial Complex · SVM · Ontology

## 1 Introduction

Nowadays, people share their ideas, feelings, and reactions via social networks to everyday-life events, politics, news, social issues, music, etc. Twitter seems to be the preferred popular social networks to express opinions and feelings about any kind of event, through tweets, generally composed of short sentences. There is a great interest in investigating human behavior in social networks, especially analyzing tweets, with the purpose to capture feelings, emotions expressed in natural language and study reactions about that specific event. Tweet analysis finds applications in many fields, such as business [7], marketing [3], political consensus analyses [10] and more.

Sentiment analysis has been widely used to interpret the sentiment polarity (positive, negative and neutral) from social media [8,11] and web documents [6]. Text mining [13] and Natural Language Processing (NLP) [12] have been widely used to extract feelings in the language structure, focusing on developing new text preprocessing methods including tokenization, stemming and other

techniques to improve sentiment classification. The lexical analysis alone can not be enough to depict sentiments and opinions, because tweets and posts are written in colloquial language and strongly depend on context [1, 4, 5]. Therefore, tweet comprehension requires not only the analysis of the document and the paragraph, but also the sentence, clause and concept analysis. In fact, some trends in literature are aimed at concept extraction to better interpret text meaning and categorize tweets [4].

A thorough interpretation of the emotions and opinions from tweets requires knowledge from different fields, such as linguistic, psychology, cognitive science, sociology and ethics. To this purpose, Sentic computing is a multi-disciplinary approach which bridges computer as well as social sciences to better recognize and depict opinions and sentiments over the Web [9]. Therefore, affective ontologies and emotion-specific reasoner tools exploit social sciences knowledge for the opinion and sentiment interpretation [9], as well as the collective emotions influence human behavior [2].

In addition to emotion extraction difficulties, the main issue about tweet analysis is due to the continuous tweet stream, coming from people of different cultures, over all the world, expressing their opinions, sentiments on different topics [5]. The use of hashtags in tweets supports the detection of user behavior by the tweet trend analysis [8, 13].

This paper presents an emotion-based classification system. The system also achieves a linguistic topological space of emotion-based concepts. These concepts characterize the emotional feelings expressed in a trend (associated with one or more hashtags) about an event, and supports the training of the multi-class Support Vector Machine (SVM) classifier to identify emotions expressed in tweets. The output of classifier is relaxed: not just the dominant feeling detected on the event, but nuances of emotions are returned, thanks to the intensity degree calculated by the classifier result. In order to improve the feeling analysis, an emotional concept ontology is also built and queried to provide meaningful views on the emotional reactions to events. For instance, let us consider the hashtag *#astarishborn*, the proposed approach depicts people reactions to the film in terms of feelings, such as love, open, angry, sad, etc.

The paper is organized as follows: Sect. 2 provides an overview of the system, while Sect. 3 discusses tweet collection and data preprocessing. Section 4 presents the emotional concept extraction from tweets. Then, Sects. 5 and 6, respectively, present soft classification in emotional classes and the ontology to model the collected data to return emotion-based views. System tests are presented in Sect. 7, then Sect. 8 concludes the paper.

## 2 Model Overview

Figure 1 shows the logical overview of the system as a data-flow through its main components. The collection of tweets is given as input to the *Tweet Analysis* component, that arranges the tweets into hashtag-based documents, to facilitate their processing. The component parses the document text by using the Natural

Language Processing (NLP) activities, to select emotional terms; these terms will form the feature space of the document-term (DT) matrix. The matrix is given as input to the *Emotional concept extraction* component, in charge of building a topological space of emotional terms. The Simplicial Complex method [4] is used to discover the strong correlation among terms in the multidimensional space. It generates emotion-based concepts with different term aggregation degrees: basic concepts (BCs) and extended concepts (CETs). The extracted BCs and CETs are provided to *Emotion classification* and *Ontology-based emotion view* components. *Emotion classification* component discovers the emotional BCs that are relevant to the hashtag-based documents. The found BCs are used to annotate the DT matrix for the classifier training. The *Ontology-based emotion view* analyses relations among BCs and CETs to automatically enrich the ontology model with new instances of the discovered emotional concepts. SPARQL queries on this ontology provide better insights on human behavior when a specific event described by hashtag is sought.

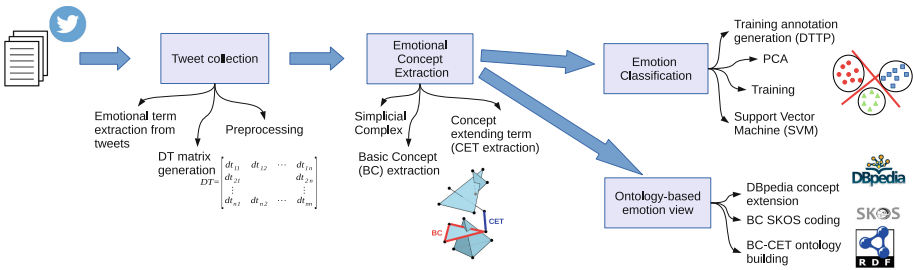


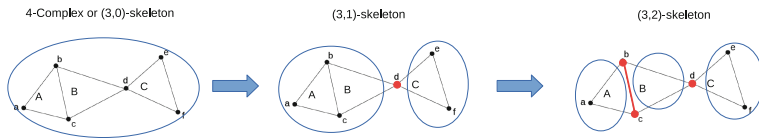
Fig. 1. System architecture.

### 3 Tweet Analysis

The tweet stream is analyzed and arranged in order to produce a document associated with the same hashtag. Thus, each document represents a social tweet trend associated with an event, described by a specific hashtag. The document collection is then parsed by NLP techniques in order to discard stop words, numbers, and then the remaining words are stemmed and further filtered according to a set of emotional term collection. The emotional terms are single terms that can be nouns, adjectives and adverbs expressing an emotional judgement about something, such as “thankful”, “admiration”, etc. or a particular mood triggered by something (i.e. “satisfied”, “sad”, “delighted”). Thus, the meaning can be positive as well as negative. Often, these terms are not meaningful of themselves, sometimes they need interpretation. For instance, “Good job” could not always be a compliment, in sarcastic sentences it could also be a negative remark about the action of somebody (i.e., “Good Job, Einstein!!!!”). These terms are, generally, related to specific classes of emotions. For instance, “devoted”, “passionate”

refer to love-related emotions, while terms such as “annoyed”, “irritated” refer to angry-related emotions. The emotional terms are individual or atomic words, that in the natural language, are associated with a feeling, although they are not often enough to fully interpret the emotion expressed. In this approach, an Emotional Term List (ETL), describing a list of typical emotional terms<sup>1</sup>, has been used. It is composed of sixteen emotional classes (also called *ETL emotional classes*), each one, in turn, composed of a set of words that are synonyms or words with similar meaning; a half of them include words expressing pleasant feelings (“open”, “happy”, “alive”, “good”, “love”, “interest”, “positive”, “strong”), whereas the remaining classes are related to unpleasant feelings (“angry”, “depressed”, “confused”, “helpless”, “indifferent”, “afraid”, “hurt”, “sad”).

Terms in the ETL file are employed to detect and select the emotional terms in each (hashtag-based) document of tweets. The document-term matrix is indeed built considering all the documents, given the (stemmed) terms in the ETL. The cell  $(i, j)$  of matrix represents the importance of the emotional term  $i$  in the document  $j$ , and is valued by the TF-IDF (term frequency-inverse document frequency) metric. The metric offsets the frequency of a term in a document with the frequency of the same term in the whole corpus. The rationale behind the metric is to reduce the importance of the most frequent terms in general, which are often little significant.

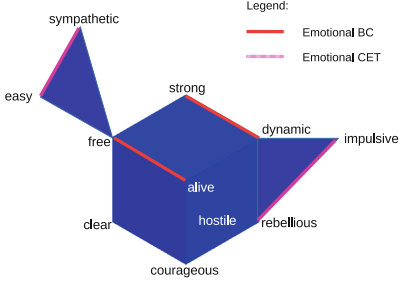


**Fig. 2.** The simplicial complex decomposition into simplexes A, B and C through skeleton definition

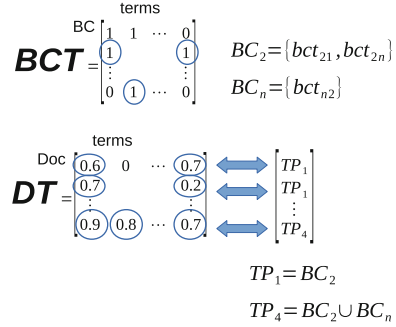
## 4 Emotional Concept Extraction via Simplicial Complex

The emotional terms are atomic terms, that can have emotional meaning, even though the proper meaning is often related to the context (surrounding words that make more explicit the intended sense) where the term appears. Understanding the actual sentiment expressed in the written language is not trivial, especially in tweets, whose reduced text length amplifies the difficulties in capturing sarcastic and ambiguous sentences. Thus, the same term could also assume different meanings depending on the context/situation. Our idea is to delineate a kind of context, by relating terms expressing emotions with other terms in sentences, with the purpose to outline an emotional concept accurately. The

<sup>1</sup> <http://www.psychpage.com/learning/library/assess/feelings.html>.



**Fig. 3.** BCs and CETs from simplicial complex structure



**Fig. 4.** DT and BCT matrices

*Emotional Concept Extraction* component is in charge of building the emotional concepts. It achieves the simplicial complex model, that is a geometric structure composed of several geometric figures, such as points, lines, triangles, squares, etc. More specifically, the simplicial complex is a finite collection of simpler geometric structures, called simplexes. Simplexes are a convex hull of  $n + 1$  independent vertices (i.e., 2-Simplex is a triangle). Simplexes could also be formed of sub-structures, called faces (i.e., triangles are composed of lines and points). The Simplicial Complex not only contains its simplexes, but also the faces connecting them. The connected simplexes in a Simplicial Complex could be separated by gradually removing faces connecting them at each step. This way, different complex skeletons are generated. Figure 2 shows this progressive decomposition of the initial simplicial complex. The two simplexes  $A$  and  $B$  are connected by the segment  $(b, c)$ , while  $B$  and  $C$  are connected by the point  $(d)$ . Firstly, the lowest dimension faces are removed, thus, points are removed generating the  $(3,1)$ -skeleton,  $d$  is removed and  $B$  and  $C$  are not yet connected. Then, the segments are removed,  $(a, b)$  included. Therefore, the  $(3,2)$ -skeleton is formed and  $A$  and  $B$  are no longer connected. This process is used to analyze the topological space (i.e., the whole simplicial complex), through its skeletons which represent different conceptualization levels. Since each point is a term, the simplicial complex of terms at skeleton  $k$  represents a linguistic topological space at the  $k^{th}$  level of detail. The simplicial complex has been proved to be useful to analyze linguistic topological spaces and extract context-related conceptualization, called Basic Concepts (BCs) and Concept Extending Terms (CETs) [4]. BCs represent specific emotional concepts, composed as a simplexes of highly-related emotional terms. The CETs, extending the BC, add more general emotional terms to the BC, providing more complex context-based conceptualization. Figure 3 shows emotional BCs and CETs built on emotional words extracted from a corpus of documents. The most related terms in the text are grouped determining the emotional BCs, such as (“free”, “alive”), (“strong”, “dynamic”). These BCs represent the basic feelings expressed in the text. The CET (“sympathetic”, “easy”)

better explains BC (“free”, “alive”), while the CET (“impulsive”, “rebellious”) better describes BC (“strong”, “dynamic”). Emotional BCs represent the emotional classes used in the training phase of our SVM classifier. The extracted BCs are arranged in a Basic Concept-emotional Term matrix (BCT). This matrix has the BCs, on the rows, and the emotional terms in the BCs, on the columns. A cell value  $(i, j)$  in BCT is 1 if the  $j$ -th emotional term belongs to the  $i$ -th BC, 0 otherwise. Let us notice that the emotional terms in each BCT are the same in the feature set of DT. An example of BCT definition is shown in Fig. 4.

## 5 Emotion-Driven Classification

This section presents the classification of the hashtag documents. The three following subsections present the training setting, the cross validation and the soft classification of the classifier outputs, respectively.

### 5.1 Target Class Generation by Emotional Conceptualizations

Since no prefixed emotion-based classes are available for the hashtag-based tweet stream, the emotional concepts (BCs), built via the simplicial complex, provide a way to compute the appropriate emotional class. Recall that each emotional class described in Sect. 3 is composed of emotional terms present in ETL; for instance, the ETL class identified by label *Happy* is composed of words like *ecstatic*, *joyous*, *gleeful*, etc. Since ETL is the feature set of the two matrices DT and BCT, the predominant classes to associate with each document are sought, taking the emotional terms in ETL in account. Specifically, the *Emotion-based Classification* component selects the BC terms that are also in a hashtag-based document. This way, a list of all the BCs (sharing terms with DT matrix) for each document is generated. The selected BCs, namely Hashtag-related BCs, describe the emotions related to the tweet trend, described by that hashtag (document). More formally,

**Hashtag-Related BCs of a Document.** Let  $BC = \{x_1, x_2, \dots, x_n\}$  be a basic concept (BC) and  $H = \{z_1, z_2, \dots, z_n\}$  a hashtag-based document,  $BC$  is related to  $H$  if and only if each emotional term  $t$  ( $t \in \text{ETL}$ ) in  $BC$  is also in  $H$ :

$$BC \subset H \leftrightarrow \forall t \in BC, t \in H \leftrightarrow x_t = 1 \text{ and } z_t > 0 \quad (1)$$

where  $x_t$  and  $z_t$  are the values assumed by the term  $t$  in BC and H (a row of DT), respectively.

All the *Hashtag-related BCs* associated with a hashtag document form an *emotional topic*, which is composed of all the emotions expressed by the discovered concepts in that hashtag stream.

**Emotional Topic of a Document.** Let  $H$  be a hashtag document and  $BC$  the hashtag-related BC to  $H$ , the emotional topic  $T_H$  associated with the document  $H$  is defined as the union of each  $BC$  related to  $H$ :

$$T_H = \bigcup_{\forall BC \subset H} BC \quad (2)$$

The emotional topic is composed of all the terms in the union of all the *Hashtag-related BCs* of a document  $H$ . The topic enables the ETL class candidate selection to get the final annotation class for that hashtag document.

**Hashtag Annotation Class.** An ETL emotional class  $C_i$  of an emotional topic  $T_H$  is assigned to the hashtag document  $H$  if  $|C_i \cap T| \geq |C_j \cap T|, \forall j = 1, \dots, n$ , (where  $n$  is the number of ETL emotional classes).

In other words, ETL class containing the largest common subset of terms (in BCs) with the topic  $T_H$ , associated to the document  $H$ , will be taken as the target class for  $H$ . The DT matrix with the class annotation is used to train the classifier.

## 5.2 Preprocessing, Cross Validation and Training

Our system employs Support Vector Machine (SVM) to classify the hashtag documents in ETL emotional classes.

As first step, the training and the test sets have been formed by dividing the annotated DT matrix in the 80% for the training and the 20% for the test. The DT division has been done by keeping the proportions among the classes. Then, the SVM parameters have been estimated by employing the k-fold method with  $k = 5$ . The high number of variables generally decreases classification performances and accuracy. In this case, the Principal Components Analysis (PCA) is applied to the dataset to detect the linearly independent variables at maximum variance. The number of the components to select is detected by cross-validation. The data reduced with PCA are also standardized in order to improve SVM accuracy. After cross validation, PCA-based data reduction and standardization have been performed, the model is trained.

## 5.3 Soft Classification

The human mood is often a blending of different emotion, with different intensity degrees. Sometimes, it is not possible to straightly label a feeling because it is come a mix of sensations, sentiments, there are not easy to recognize. Bearing this in mind, our system achieves a soft classification of the hashtags to better represent people’s emotions and reactions to events, that in this case, are described by the relative hashtag trends. The multi-class SVM method employs the pairwise classification, which builds a classifier for each pair of classes. Therefore, given  $m$  classes,  $m \cdot (m - 1)$  classifiers are employed. Each classifier chooses one class from the pair, which is assigned to. In a hard classification scheme, the most voted class on all the pairs is chosen to classify the hashtag (majority voting algorithm).

Our system, instead, does not only take the most voted class, but also the votes for each class over the  $m \cdot (m - 1)$  pairs. Each vote number associated to each ETL emotional class normalized in the range  $[0, 1]$ , represents the intensity of that feeling. The degree of the feeling intensity allows a soft classification of the hashtag to ETL classes, which thoroughly expresses the wider nuances of

people’s mood to something (i.e., hashtag associated with an event, topic, etc.). An example of a classification of a tweet trend: (open 0.0), (happy 0.5), (alive 0.1), (good 0.0), (love 0.4), (interest 0.0), (positive 0.1), (strong 0.0), (angry 0.0), (depressed 0.0), (confused 0.0), (helpless 0.0), (indifferent 0.0), (afraid 0.0), (hurt 0.0), (sad 0.0). The value associated with each ETL emotional class describes the intensity degree of that emotion expressed in the tweet trend with a given hashtag. It is evident that the comprehensive feeling is pleasant since terms in the predominant emotional classes such as *happy*, *love* were discovered in the tweets with that hashtag. Also *alive* and *positive* contribute to completely depict the nuances of the human mood. The soft classification allows more accuracy in the description of the feeling, with respect to the crisp classification that would have expressed only happy emotion as people’s mood. The soft classification better depicts all the people’s feeling and emotional reaction to the hashtag/event.

## 6 Ontology-Based Emotion View

A semantic model encodes all the given and generated concepts, participating to the system pipeline shown in Fig. 1. The BCs and CETs are employed to build a concept ontology on the emotional concepts (Emotional Concept ontology). One of the first task of the *Ontology-based emotion view* component is indeed to link each generated BC to DBpedia<sup>2</sup>, which is an online semantic dataset to extract structured content of Wikipedia resources, in order to enrich our ontology with the corresponding well defined DBpedia concepts. The modeling of BC is based on the *concept* class of SKOS<sup>3</sup> ontology. SKOS gives specifications and standards to support the use of Knowledge Organization Systems (i.e. thesauri, classification schemes, and taxonomies). It provides a knowledge model to represent and classify high-level concepts and the semantic relations among them. The concept of relations expresses how concepts are related to each other. Each BC is coded as a SKOS concept, and related to the other BCs defining RDF and SKOS properties. Three properties have been used to model the semantic relationships among the BCs: *skos:related*, *skos:narrower* and *rdfs:seeAlso*. These properties allow building a concept hierarchy based on relations among the BCs, and extend the BCs with the CETs and the DBpedia matching resources. The *skos:narrower* property is employed to model the relation between the hashtag document and its individual emotional topic (see Eq. (2)), which is in turn, the union of all the BCs present in the hashtag document. A hierarchy is built by the narrower property, that relates the hashtag documents to the emotional topics and their BCs. The *rdfs:seeAlso* property is employed to relate each BC (SKOS concept) to the DBpedia resource expressing the homonymous emotional concept. The DBpedia resource enriches the BC with additional information (i.e., definitions, related concepts, contexts in which BC is present).

Since each BC is also related to the CETs (that better contextualize the BCs), the *skos:related* property associates BC with CETs. This way, all the resources

<sup>2</sup> <https://wiki.dbpedia.org/>.

<sup>3</sup> <https://www.w3.org/2004/02/skos/>.



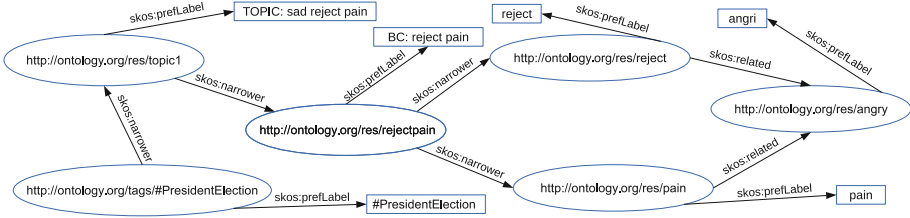


Fig. 5. Emotional concept ontology extract.

linked to a BC through *skos:related* provide a contextualized information about the BCs. The described ontology model is the basic skeleton of our emotional knowledge. The result from the classification provides a dynamic component of the ontology, viz., the population, composed of the actual terms, BCs, CETs, tweet trends, and the extracted emotional valences. Figure 5 shows an extract of the Emotional Concept ontology on (the documents generated by) the trend with the hashtag `#PresidentElection`. This hashtag is related to its emotional topic, labeled *topic1*, composed of two BCs: *sad* and *rejectpain*. The *rejectpain* BC is in turn, related to its terms: *reject* and *pain* terms. DBpedia resources, if present, are also related to the BCs; for instance, the BC *sad* is related to *Sadness* through *rdfs:seeAlso* property. BCs could also be related to their eventual CETs, such as *reject* and *pain* which are related to their common CET *angry*.

SPARQL queries can generate views on the knowledge base, that provide a better insight on the hashtag emotion classification, and better interpretations of the emotion-based human behavior in response to some events and tweet topics, through the continuous tweet streams. The ontology provides a simple representation of the emotional concepts, anyway, more refined semantics will be explored in the future.

## 7 Experimental Results

Tests have been conducted on tweets, selected according to hashtag in the top trends of the moment, from June to August, 2018. Hashtags have been extracted considering generic events related to week days (i.e. `#WednesdayWisdom`, `#ThursdayThoughts`, `#TBT`, `#FridayFeeling`) and important international events, such as `#PrimaryElection`, `#PMQs`, as well as to more frivolous events `#NationalBestFriendDay`, `#NationalChocolateIceCreamDay`. Many hashtags refer to showbiz, music, sport and TV events, such as `#TonyAwards`, `#WMA`, `#loveisland`, `#astarisborn`, `#GNTM`, others to politics, social issues, news or religious events, such as `#Trump`, `#politicalcorrectness`, `#RSSTritiyaVarsh`, or even some initiatives or campaigns (i.e. `#FlirtWithYourCity`, `#GlobalRunningDay`).

Tweets related to the same hashtag have been collected and saved in a document. Tweets have been stored on a server employed for tests, which is equipped with 32 GB ram, 2 TB SATA HDD and 8 quadcores, that have speed up the

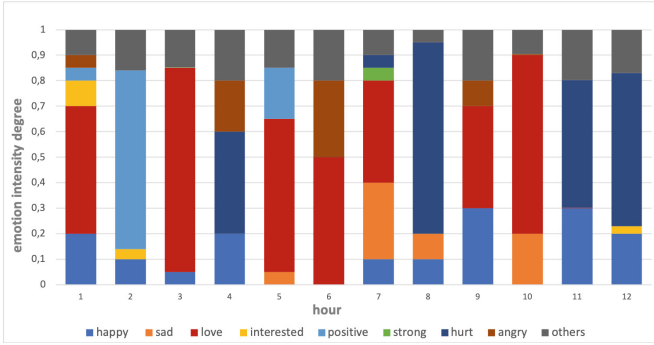
**Table 1.** Experimental results by varying configuration setting. Each row shows the experiment (Exp), document number (doc), feature number (Feats), pleasant (plF) and unpleasant features (unF), pleasant (plC) and unpleasant categories (unC), Accuracy (Acc)

Exp	Doc	Feats	plF	unF	plC	unC	Acc
1	5617	250	126	124	7	6	0.57
2	5617	75	38	37	6	5	0.58
3	5617	250	128	122	6	5	0.62
4	5061	100	100	0	6	0	<b>0.70</b>
5	5061	109	58	51	6	6	0.55
6	4000	105	54	51	6	5	<b>0.7</b>
7	4000	70	36	34	5	4	<b>0.7</b>
8	4000	46	16	30	2	3	<b>0.81</b>
9	4000	30	11	19	6	2	<b>0.89</b>
10	4000	30	11	19	2	3	<b>0.97</b>
11	3000	150	150	0	6	0	<b>0.68</b>
12	2000	30	16	14	6	4	<b>0.90</b>
13	2000	150	64	86	6	7	0.55

tweet collection and the execution of the overall system. The hashtag corpus contains 5617 documents, each one containing tweets with a specific hashtag. Each document is composed of about 2000 tweets, hence the dataset includes around 11,234,000 tweets.

Table 1 shows the results of the experimentation, we conducted, by varying the number of documents, features, and accordingly, the pleasant and unpleasant classes. Some classes have been discarded because their terms had a few occurrences in the document corpus. The experimentation shows the accuracy, by progressively varying some values. On average, the pleasant classes have more occurrences than unpleasant ones, due mainly to the tweet trend content, generally expressing positive emotions. Experiments namely *Exp 4* and *Exp 11* reveal better classification results when there are no unpleasant features in the corresponding classes. Most of emotional terms from the unpleasant classes are often rare in the tweet collection, indeed, on the whole document collection, unpleasant classes of features badly impact on accuracy (see *Exp 5*) as well as there are some ambiguous terms, e.g., *open* and *alive* related to some pleasant classes that could not refer to emotions. Some experiments provide considerable improvement in the classification results removing these terms, reducing consequently the emotional classes (*Exp 10*).

Additional experiments evidence meaningful performance rate of the classifiers after using PCA to reduce the number of features. For instance, by reducing the number of features from *Exp 6* to *Exp 9*, but keeping the same number of documents, the accuracy increases significantly. By increasing the number of



**Fig. 6.** Emotions associated with the hashtag #WorldCup2018 at each hour from the first appearance of the hashtag

documents as well, the selected features maintain high values of accuracy. Experiments *Exp 12* and *Exp 9* indeed, achieve an accuracy around the 90% passing through 2000 documents to 4000 documents, respectively.

The emotion analysis conducted via tweets also allows the evaluation of the reaction to events over time. Figure 6 shows people’s reactions to the World Cup 2018, as emotional states on tweets with the hashtag #WorldCup2018. The tweets have been collected from streaming at each hour from the first appearance of the hashtag. The trend of the emotional reactions results from the hashtag-based document classification at each hour, which describes the evolution of the people feeling during the event time. Noticed that the *Love* and *Hurt* sentiments alternate as time goes by. This trend describes supporters’ reactions to team matches, expressing positive feelings (i.e., *Love*) if their team won or negative (i.e., *Hurt*) if it lost. The stacked columns, present in figure, represent the soft emotional classification for the hashtag at each hour. The wider column portion at the hour  $h$  describes the highest ranked emotional class at that moment. The soft classification provides interesting insights on the reactions to the event. For instance, at some hour, even though the highest ranked feeling is pleasant, the second highest ranked one is unpleasant (e.g., *Love* and *Angry* at hour 6) and viceversa. In many cases, the first two highest ranked sentiments have also very close ranking values. This trend is quite common at several hours, such as *Love* and *Angry* at hour 6, *Love* and *Interested* at hour 7, *Love* and *Happy* at hour 9, etc.

## 8 Conclusion

The paper presented an emotion-based classification approach from stored tweet stream. Emotional terms are placed in a linguistic topological space that supports the detection of emotional concepts, used to train a Support Vector Machine (SVM) classifier. Final result is a “soft” classification that provides a complete description of people’s reactions to the events represented by the

hashtags over time. In addition, an emotional concept ontology provides views on the extracted emotional concepts and better support the analysis of people's behavior and reactions. Future works will further investigate the emotional concept extraction for mixed emotion detection.

## References

1. AL-Sharuee, M.T., Liu, F., Pratama, M.: Sentiment analysis: an automatic contextual analysis and ensemble clustering approach and comparison. *Data Knowl. Eng.* **115**, 194–213 (2018). <https://doi.org/10.1016/j.datak.2018.04.001>
2. Alharbi, A.S.M., de Doncker, E.: Twitter sentiment analysis with a deep neural network: an enhanced approach using user behavioral information. *Cogn. Syst. Res.* **54**, 50–61 (2019). <https://doi.org/10.1016/j.cogsys.2018.10.001>
3. Çali, S., Balaman, Ş.Y.: Improved decisions for marketing, supply and purchasing: mining big data through an integration of sentiment analysis and intuitionistic fuzzy multi criteria assessment. *Comput. Ind. Eng.* **129**, 315–332 (2019). <https://doi.org/10.1016/j.cie.2019.01.051>
4. Cavaliere, D., Senatore, S., Loia, V.: Context-aware profiling of concepts from a semantic topological space. *Knowl.-Based Syst.* **130**, 102–115 (2017). <https://doi.org/10.1016/j.knosys.2017.05.008>
5. Cotelo, J., Cruz, F., Enríquez, F., Troyano, J.: Tweet categorization by combining content and structural knowledge. *Inf. Fus.* **31**, 54–64 (2016). <https://doi.org/10.1016/j.inffus.2016.01.002>
6. Hussein, D.M.E.D.M.: A survey on sentiment analysis challenges. *J. King Saud Univ.-Eng. Sci.* **30**(4), 330–338 (2018). <https://doi.org/10.1016/j.jksues.2016.04.002>
7. Li, X., Wu, C., Mai, F.: The effect of online reviews on product sales: a joint sentiment-topic analysis. *Inf. Manag.* **56**(2), 172–184 (2019). <https://doi.org/10.1016/j.im.2018.04.007>, social Commerce and Social Media: Behaviors in the New Service Economy
8. Öztürk, N., Ayvaz, S.: Sentiment analysis on Twitter: a text mining approach to the Syrian refugee crisis. *Telemat. Inform.* **35**(1), 136–147 (2018). <https://doi.org/10.1016/j.tele.2017.10.006>
9. Poria, S., Cambria, E., Winterstein, G., Huang, G.B.: Sentic patterns: dependency-based rules for concept-level sentiment analysis. *Knowl.-Based Syst.* **69**, 45–63 (2014). <https://doi.org/10.1016/j.knosys.2014.05.005>
10. Yaqub, U., Chun, S.A., Atluri, V., Vaidya, J.: Analysis of political discourse on Twitter in the context of the 2016 US presidential elections. *Gov. Inf. Q.* **34**(4), 613–626 (2017). <https://doi.org/10.1016/j.giq.2017.11.001>
11. Yoo, S., Song, J., Jeong, O.: Social media contents based sentiment analysis and prediction system. *Expert Syst. Appl.* **105**, 102–111 (2018). <https://doi.org/10.1016/j.eswa.2018.03.055>
12. Zeroual, I., Lakhouaja, A.: Data science in light of natural language processing: an overview. *Procedia Comput. Sci.* **127**, 82–91 (2018). <https://doi.org/10.1016/j.procs.2018.01.101>. Proceedings of the first international conference on intelligent computing in data sciences, ICDS 2017
13. Zhang, F., Fleyeh, H., Wang, X., Lu, M.: Construction site accident analysis using text mining and natural language processing techniques. *Autom. Constr.* **99**, 238–248 (2019). <https://doi.org/10.1016/j.autcon.2018.12.016>