An ontology-driven perspective on the emotional human reactions to social events

Danilo Cavaliere* and Sabrina Senatore

Department of Information and
Electrical Engineering and Applied Mathematics,
University of Salerno,
Fisciano (Sa), Italy
Email: dcavaliere@unisa.it
Email: ssenatore@unisa.it
*Corresponding author

Abstract: Social media has become a fulcrum for sharing information on everyday-life events: people, companies, and organisations express opinions about new products, political and social situations, football matches, and concerts. The recognition of feelings and reactions to events from social networks requires dealing with great amounts of data streams, especially for tweets, to investigate the main sentiments and opinions that justify some reactions. This paper presents an emotion-based classification model to extract feelings from tweets related to an event or a trend, described by a hashtag, and build an emotional concept ontology to study human reactions to events in a context. From the tweet analysis, terms expressing a feeling are selected to build a topological space of emotion-based concepts. The extracted concepts serve to train a multi-class SVM classifier that is used to perform soft classification aimed at identifying the emotional reactions towards events. Then, an ontology allows arranging classification results, enriched with additional DBpedia concepts. SPARQL queries on the final knowledge base provide specific insights to explain people's reactions towards events. Practical case studies and test results demonstrate the applicability and potential of the approach.

Keywords: emotion soft classification; SVM; emotional concept ontology; RDF; SKOS; SPARQL queries; sentiment analysis; tweets; simplicial complex; emotional concept extraction.

Reference to this paper should be made as follows: Cavaliere, D. and Senatore, S. (2021) 'An ontology-driven perspective on the emotional human reactions to social events', *Int. J. Metadata Semantics and Ontologies*, Vol. 15, No. 1, pp.23–38.

Biographical notes: Danilo Cavaliere received his Master degree in Computer Science in 2014 and PhD degree in 2020 from the University of Salerno, Italy. He reviewed papers for various journals, including *Information Sciences*, *Expert Systems with Application*, *IEEE Transactions on Cognitive Communications and Networking* and *IEEE Access*, he was in the Program Committee for *IEEE Symposium on Intelligent and Robotic Agents (IEEE SSCI'19)* and *IEEE Symposium on Intelligent Agents (IEEE SSCI'20)*. His research interests are in the areas of artificial and computational intelligence, data science, data mining and knowledge discovery, areas in which he has published many papers.

Sabrina Senatore is an Associate Professor of Computer Science at the University of Salerno, Italy. She is the chair of *IEEE CIS Task Force on Intelligent Agents (TFIA)*, an Associate Editor of *Information Science*, *Elsevier* and *IEEE Transactions on Cybernetics*; she is also an Editorial Board Member of the *Applied Intelligence Journal* and *International Journal of Computational Intelligence Theory and Practice*. Her current research interests include the development and application of intelligent systems based on the combination of techniques from soft computing, computational intelligence, text mining, information retrieval, semantic web and software agents, areas in which she has published extensively.

This paper is a revised and extended version of a paper entitled 'Emotional concept extraction through ontology-enhanced classification' presented at the '13th International Conference Metadata and Semantic Research (MTSR'19)'.

1 Introduction

Nowadays, people use social networks to keep in touch with friends, look for information about products, share their thoughts, feelings and reactions to many everyday-life events, related to politics, news, social issues, films, music, etc. Among the most popular social networks, Twitter seems to be the one preferred by people to express their opinions about any kind of event. There is a great interest in investigating human behaviours from social networks, due to the high number of Twitter users all around the world commenting relevant events. Twitter users often become aware of events by directly reading tweets, and they react accordingly by writing their thoughts through tweets or sharing tweets written by other users through re-tweeting. Tweet analysis finds applications in many fields, such as business (Li et al., 2019a), marketing (Çal and Balaman, 2019), political consensus analyses (Yaqub et al., 2017) and more. Emotion extraction by tweet analysis may support various applications, such as the target selection for the launch of a product or voting intention analysis. The extraction of people's emotional reactions from tweets is not an easy task due to differences among people and communities who write tweets and tweet features. Some tweet features may complicate the reaction interpretation. Tweets are generally composed of short sentences written in colloquial language, which require knowledge about the dialect, slang and idiomatic expressions used by the tweet writer.

Furthermore, tweets reflect the people's moods, which are a combination of feelings, perceptions, the interpretation of which could be quite complex, especially considering that people reaction is strongly affected by their nationality, community, culture and education. Therefore, emotional reaction recognition requires to put users' reactions into a context, that can also serve the analysis of impacts and relevance of reactions. Approaches for reaction extraction are generally aimed at discovering the dominant reaction common to the highest number of people, in order to identify typical behaviours towards a product or event. For instance, in marketing one-to-one applications, there is the need to analyse human behaviour to satisfy all the tastes and requirements as much as possible.

The work introduced in Cavaliere and Senatore (2019) collects and classifies tweets to detect people's emotional reactions towards events in the form of emotional classes, such as love, open, angry, sad, etc. The approach also introduced a preliminary ontology model for modelling feelings expressed in tweets. This paper presents an extension of the original framework presented in Cavaliere and Senatore (2019), by re-designing and upgrading the preliminary ontology model to integrate the classification results with the emotional concepts detected via a linguistic topological space of the terms extracted from tweets. The novel ontology model allows a better explanation of people's emotional reactions through query-based views. For instance, let us consider the hashtag *#astarisborn*, the proposed approach depicts people's reactions to the film in terms of feelings, such as happy, interested, angry, afraid, etc. Then, these reactions can be better explained by querying the generated ontological knowledge base to extract emotional concepts and terms (i.e., classification result: angry, query-based explaining terms: rancorous, resentful).

The approach contribution is manifold.

- Emotional concepts are extracted from tweets by using a linguistic topological space of terms, built through the geometrical structure called simplicial complex.
- People's emotional reactions are detected and described through a soft classification based on Support Vector Machine (SVM), trained on the extracted emotional concepts.
- An emotional concept ontology is introduced to model and integrate knowledge on the emotional concepts, extracted thanks to the simplicial complex, and people's emotional reactions detected through classification.
- Queries are presented to generate specific views on feelings and events to detail emotional reactions to an event/hashtag and relate events according to people's reactions.

The paper is organised as follows: Related work discussion is conducted in Section 2, hence Section 3 provides an overview of the system, while Sub-section 3.1 discusses tweet collection and data pre-processing. Sub-section 3.2 presents the emotional concept extraction from tweets. Then, Sub-sections 3.3 and 3.4, respectively, present soft classification in emotional classes and the ontology to model the collected data to return emotion-based views. System is evaluated through tests and case studies in Section 4, then Section 5 concludes the paper.

2 Related work

Sentiment analysis has been widely studied in many application domains, such as social media contents (Yoo et al., 2018; Öztürk and Ayvaz, 2018) and web documents (Hussein, 2018) by interpreting the sentiment among positive, negative and neutral classes (text polarity classification). To this purpose, many approaches, such as the one proposed in Basha and Rajput (2018), present supervised methods to predict sentiments on tweeter documents. The role of Text Mining and Natural Language Processing (NLP) is crucial to support emotion extraction by analysing the natural language structure. Symeonidis et al. (2018) and Singh and Kumari (2016) provided solutions to improve the lexical analysis through NLP methods including tokenisation, stemming, parsing and stop word removal. Notwithstanding the benefits introduced, the lexical analysis is not enough to depict feelings and reactions, due to the colloquial language in social content, i.e., people generally use slang and ironic sentences in tweets that strongly depend on the education, culture and context (AL-Sharuee et al., 2018; Cavaliere et al., 2017; Cotelo ett al., 2016). Consequently, a thorough comprehension of opinions and reactions expressed in tweets demands not only a lexical analysis at sentence and clause level but also at the concept level. To this purpose, some trends in literature propose solutions (Cavaliere et al., 2017) aimed at discerning robust concepts from a spatial representation of correlated terms for tweet classification.

A deep comprehension of the emotional reactions from tweets requires knowledge from different research fields, including linguistics, cognitive science, sociology, psychology and ethics. Sentic computing is a multi-disciplinary approach aimed at improving the opinion mining and sentiment extraction from web resources by integrating computer science with social sciences (Poria et al., 2014). Some trends integrate the lexical analysis with social science knowledge by using ad-hoc ontologies and reasoners to deal with opinion mining and sentiment extraction (Poria et al., 2014), as well as the analysis of human behaviours according to the collective emotions (Alharbi and De Doncker, 2019).

Beyond the difficulties related to language interpretation and emotion extraction, the main issue about tweet analysis is related to the massive data stream generated by a multitude of people participating in the Twitter community. There is a great deal of information coming from very different kinds of people all over the world, who express themselves on myriads of topics everyday (Cotelo et al., 2016).

Moreover, the use of hashtags in tweets provides gathering tweet trends for user behaviour analysis (Öztürk and Ayvaz, 2018; Zhang et al., 2019).

Starting from tweet trends, many approaches (Cotelo et al., 2016) achieve tweet categorisation employing lexical structures, such as Bag of Words (BoW) method; Ibtihel et al. (2018) used external knowledge bases, like WordNet, to enrich knowledge on tweets. The use of hashtags allows the tweet classification per topic, but does not provide a way to compare people's reactions to similar or related topics.

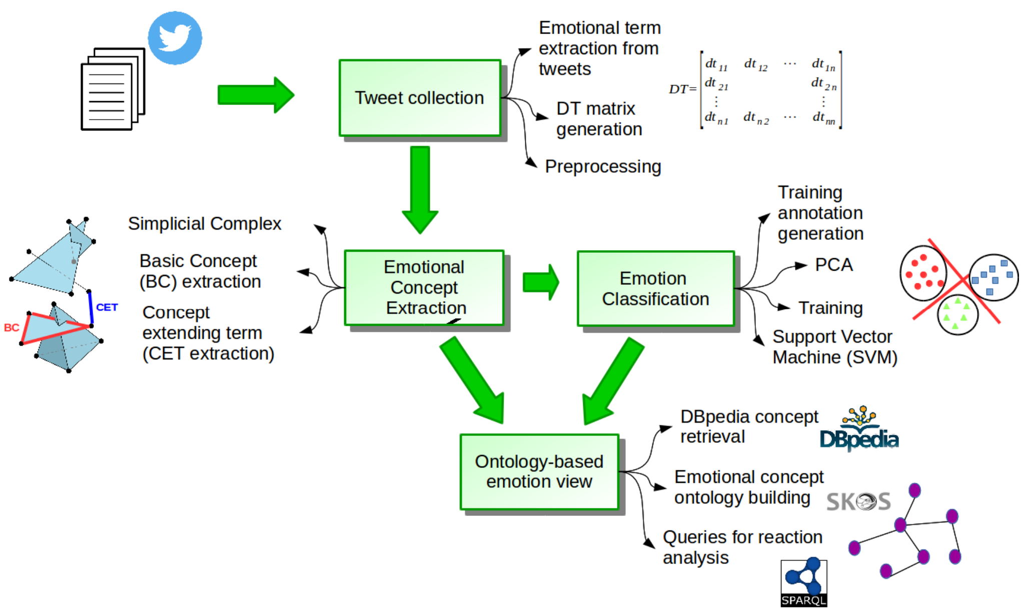
To address these issues, context-aware solutions are required to better interpret people's reactions, according to the context of the specific topic. To this purpose, ontology-based systems provide domain knowledge for improving people's behaviour analysis, such as the approach proposed in Santosh and Vardhan (2016), that introduces an ontology on product review that serves as background knowledge to learn sentiments on product features through rules. In literature, several works introduced context-aware systems to recognise people's behaviour and reactions (Li et al., 2019b; Thakor and Sasi, 2015). Li et al. (2019b) introduced a neural network model to discover the reasons behind people's reactions (emotion cause analysis) from the text. In Thakor and Sasi (2015), an ontology-based approach analyses customer dissatisfaction from tweets.

Contrary to these methods using a specific methodology, this paper introduces a context-aware approach that combines machine learning techniques and knowledge-based methods to, respectively, recognise people's reactions through classification and generate a term-based reaction explanation through ontology querying.

3 Model overview

A logical sketch of the system architecture, along with its main modules, is presented in Figure 1. As the first step, the collected tweets are provided to the *Tweet Analysis* module, which accomplishes data pre-processing. This module indeed arranges tweets into hashtag-based documents, according to the hashtag trend, and filters the so-built documents to extract emotional terms by using Natural Language Processing (NLP) techniques. The extracted emotional terms form the feature space of the Document-Term (DT) matrix. The DT matrix is given as input to the *Emotional concept extraction* module, in charge of building a topological space of the terms extracted from tweets and present in DT. To this purpose, the geometrical simplicial complex structure is built to relate the terms (Cavaliere et al., 2017). This structure allows to find correlations among the emotional terms in the multidimensional space and, accordingly, extract emotional Basic Concepts (BCs) and Concept Extending Terms (CETs). These BCs and CETs are passed to both the *Emotion classification* and *Ontology-based emotion view* modules. The former is in charge of building the label to train the emotion classifier. Precisely, it finds the BCs that are relevant to the hashtag-based documents and use them to annotate the DT matrix to train the SVM classifier. The latter analyses the relations between BCs, CETs, and the events or trends, along with the classification results, to automatically populate the emotional concept ontology with facts about the emotional reactions to the events or trends (each one of them identified with a specific hashtag), enriched with concept-related DBpedia data. Then, SPARQL queries on the knowledge base generate views for reaction analysis and explanation.

Figure 1 System architecture



3.1 Tweet collection

The first module analyses and arranges the tweet stream into groups of hashtag-based documents; in detail, a document is composed of tweets containing a specific hashtag representing the social trend associated with the hashtag and describing the specific event. The document collection is then parsed by NLP techniques to filter out numbers and stop words then the remaining words are stemmed and further filtered by comparing them with a predefined set of emotional categories, including various emotional terms. These terms are single terms expressing emotions; they can be nouns, adjectives, and adverbs expressing an emotional judgment about something, such as "admiration", "thankful", etc. or a particular mood triggered by someone or something (i.e. "sad", "satisfied", "delighted"). Term meaning can be positive as well as negative. Linguistic interpretation is definitely not trivial due to the ambiguities of natural language, sarcasm, irony, and mixed emotions. For instance, the sentence "Good job !" generally expresses praise, but, in some contexts, it can be a sarcastic comment instead, expressing a negative remark about the action committed by somebody (i.e., "Good Job, Einstein!!!!").

The emotional terms are individual or atomic words that are associated with specific feelings in the natural language, although sometimes they fail in expressing precise emotions or can be not related to emotions at all. In general, terms can be related to specific classes of emotions for instance, words such as "passionate" and "devoted" refer to love-related emotions, while "annoyed" and "irritated" are related to angry-related emotions.

In this approach, an Emotional Term List (ETL)[1] has been taken into consideration for term filtering. The list includes typical emotional terms grouped per emotional category. In detail, ETL is composed of 16 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 used to individuate and pick the emotional terms from tweets in each hashtag-based document. The feature set of the document-term matrix is composed of all the stemmed terms extracted from hashtag-document that are in the ETL. The value in the matrix cell (i, j) describes the importance of the emotional term i in the document j, calculated by the TF-IDF (term frequency-inverse document frequency) measure. The measure assesses the ratio between the frequency of a term in a document and the frequency of the same term in the whole corpus. The rationale behind this measure is to reduce the importance of the most frequent terms, which are generally not particularly relevant.

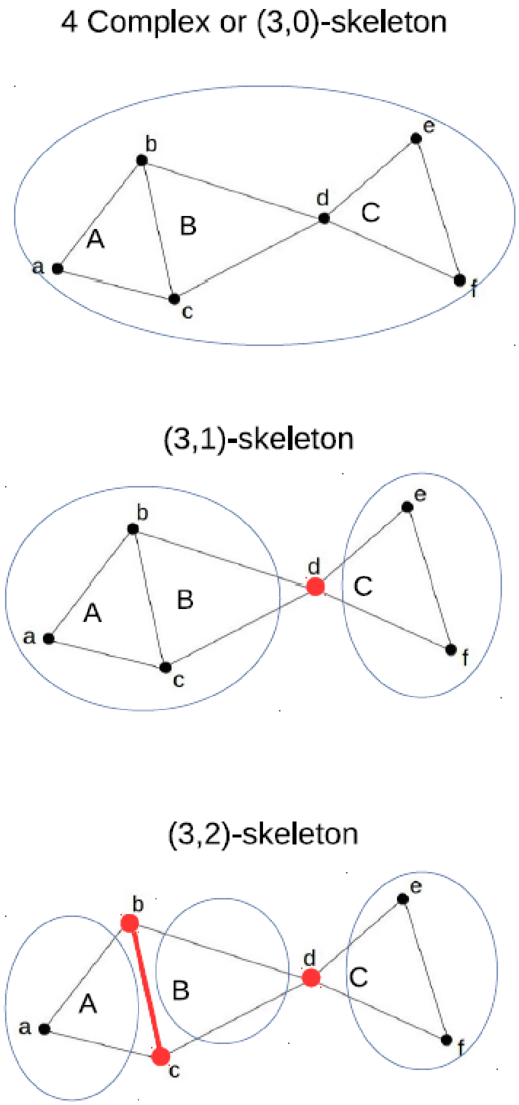
3.2 Emotional concept extraction via simplicial complex

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. The comprehension of the substantial sentiment expressed in the written language is a tough task to fulfil, especially when dealing with tweets, which generally consists of short text which makes way more difficult sentiment detection, even more in presence of sarcastic and ambiguous sentences. Thus, the only emotional term does not provide a definitive interpretation of the kind of feeling expressed, because the term may assume different meanings depending on the context/situation the term is in.

The idea at the basis of our approach is to define a context, by relating terms expressing emotions with other terms in sentences, with the aim of identifying an emotional concept clearly.

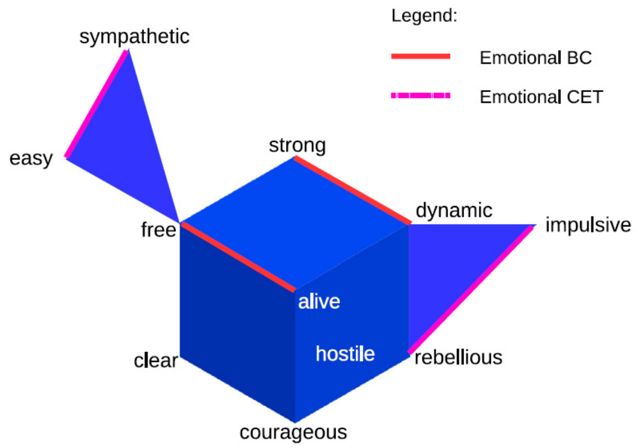
The *Emotional Concept Extraction* module is aimed at building the emotional concepts. To accomplish this task, the module employs the simplicial complex model, which is a geometric structure composed of several geometric figures of different dimensions, such as points, lines, triangles, squares, etc. More specifically, the simplicial complex is a finite collection of simpler geometric structures, called simplexes. An n-simplex represents a convex hull of $n+1$ independent vertices (i.e., 3-Simplex is a tetrahedron). Simplexes could also be composed of sub-structures, called faces (i.e., tetrahedrons are composed of triangles, lines and points). The simplicial complex is not exclusively composed of its simplexes, but it also includes their faces, including those that link distinct simplexes in the simplicial complex. The connected simplexes in a simplicial complex could be separated by gradually removing all the faces that link them in consequent steps. Through this process, different structure skeletons are built. The progressive decomposition of the initial simplicial complex into skeletons is shown in Figure 2. 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. Skeleton generation allows analysing the linguistic topological space, represented by the whole simplicial complex, through several layers (i.e., its skeletons) that represent distinct conceptualisation levels. Therefore, by considering each term as a point of the simplicial complex, the simplicial complex of terms at skeleton k represents a linguistic topological space at the *k*-th level of detail. To build the linguistic topological space over tweets, the terms extracted from tweets, are processed with NLP-based pre-processing to identify them in terms of parts of speech through POS tagging. Then, the terms, which do not belong to the ETL term list, are filtered out. The relevance of the remaining terms in each tweet document is assessed through linguistic metrics, and according to the assessed term relevance values, distances among terms are calculated to build the linguistic space over tweets through the simplicial complex (Cavaliere et al., 2017).

Figure 2 The simplicial complex decomposition into simplexes A, B and C through skeleton definition



The simplicial complex has been demonstrated to serve the analysis of the linguistic topological space and support the generation of context-related conceptualisations, called Basic Concepts (BCs) and Concept Extending Terms (CETs) (Cavaliere et al., 2017). These conceptualisations are sets composed of terms extracted from tweets and matching ETL terms. In detail, the BCs describe precise emotional concepts, represented as simplexes of highly-related emotional terms. The CETs are terms that expand the original BC by adding more general emotional terms to it, generating, this way, a more complex and context-based conceptualisation. Examples of BCs and CETs are shown in Figure 3, these concepts are generated from the simplicial complex built on the emotional terms extracted from a corpus of hashtag-based documents. The terms, that have strong relations 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. CETs, composed of terms related to the BCs, provide a further characterisation of the sentiment. For instance, the CET ("sympathetic", "easy") contributes to enrich the BC ("free", "alive"), as well as the CET ("impulsive", "rebellious") better explains the BC ("strong", "dynamic"). Emotional BCs represent emotional concepts that are related to the emotional classes in ETL to train the SVM classifier for the emotions. This process will be detailed in the next section.

Figure 3 BCs and CETs from simplicial complex structure



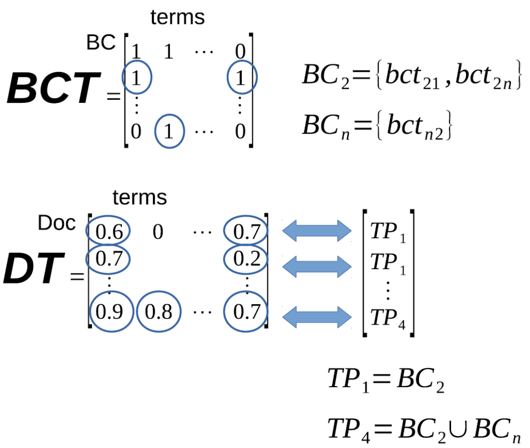
3.3 *Emotion-driven classification*

This section presents the classification of the hashtag documents. The following subsections present the training setting, the cross-validation, and the final classification, respectively.

3.3.1 *Target class generation by emotional conceptualisations*

Since no prefixed emotion-based classes are available for the hashtag-based tweet stream, the emotional concepts (BCs) provide a way to compute the appropriate emotional class. Recall that each emotional class described in Subsection 3.1 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. ETL is the feature set of the two matrices Document-Term matrix (DT), relating the documents containing tweets with their terms, and the Basic Concept-emotional Term matrix (BCT). This matrix has the BCs, on the rows, and the emotional terms in the BCs, on the columns. Examples of DT and BCT definitions are shown in Figure 4, let us notice that the emotional terms in each BCT are the same in the feature set of DT. A cell value (i, j) in DT represents that the j-th emotional term is present in the i-th BC, while a cell value (i, j) in BCT equals to 1 means that the j-th emotional term belongs to the i-th BC, it will be 0 otherwise. 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 into account. Specifically, the *Emotion-based Classification* module 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-based document.

Figure 4 Topic generation from DT and BCT matrices



More formally,

Hashtag-related BCs of a document (Cavaliere and Senatore, 2019): Let the *i*-th row of the BCT matrix $BC_i = \{x_1, x_2, ..., x_n\}$ be a Basic Concept (BC) and the *k*-th row of the DT matrix $Doc_k = \{z_1, z_2, ..., z_n\}$ a hashtag-based document, then BC_i is related to Doc_k if and only if each emotional term t ($t \in$ ETL) in BC_i is also in Doc_k:

$$BC_i \subset Doc_k \leftrightarrow \forall t \in BC_i,$$
$$t \in Doc_k \leftrightarrow x_t = 1 \; and \; z_t > 0 \tag{1}$$

where x_t and z_t are the values assumed by the term t in BC_i (the *i*-th row of BCT) and Doc_k (the *k*-th row of DT), respectively.

All the *Hashtag-related BCs* associated with a hashtag document form an *emotional topic*, which is composed of all the emotional concepts discovered from tweets on that hashtag.

Emotional Topic of a document (Cavaliere and Senatore, 2019): Let Doc_k be a hashtag document and BC_i the hashtag-related BC to Doc_k, the emotional topic T_{Doc_k} associated with the document Doc_k is defined as the union of each BC_i related to Doc_k:

$$T_{Doc_k} = \bigcup_{\forall BC_i \subset Doc_k} BC_i \tag{2}$$

The emotional topic is composed of all the terms in the union of all the *Hashtag-related BCs* of a hashtag document Doc_k.

To clarify the topic definition let us consider the example in Figure 4, that shows the definition of the emotional topic for each hashtag in DT. As it can be seen, if the location $bct[i, j]$ is 1, it implies that the *j*-th term is present in the *i*-th BC, for instance bct_{21} and bct_{2n} are present in BC_2. Then, if the same terms assume values different from 0 in DT matrix on the *k*-th row, the *k*-th hashtag document (Doc_k) will contain the BC composed of those terms, in fact, bct_{21} and bct_{2n} assume non-zero values for the first and second documents, that will be assigned with the topic TP_1 composed only of BC_2. The *n*-th row contains the terms present in BC_2 and BC_n, then the *n*-th document will be assigned with the union of the two BCs forming the topic TP_4. The topic does not represent the final annotation: it represents a summary of the emotional reactions to the hashtag. The final annotation for the hashtag document is got from an analysis of the topic and the ETL emotional classes: the ETL class, which has the most BCs in common with the hashtag topic, will be automatically chosen as the annotation for that hashtag. Since both ETL classes and BCs contain emotional terms, BCs can be subsets of ETL classes. Therefore, let us formally define the hashtag annotation class.

Hashtag annotation class (Cavaliere and Senatore, 2019): An ETL emotional class C_i of an emotional topic T_{Doc_k} is assigned to the hashtag document Doc_k if $|C_i \cap T_{Doc_k}| \geq |C_j \cap T_{Doc_k}|$, $\forall j = 1, ..., n$, (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_{Doc_k}, associated to the document Doc_k, will be taken as the target class for Doc_k. The DT matrix with the class labels is used to train the classifier.

3.3.2 Preprocessing, cross-validation and training

Once the DT matrix has been annotated, Support Vector Machine (SVM) classification is employed to assign each hashtag document with an emotional class from ETL.

In the pre-processing phase, the training and the test sets are generated from the DT matrix by selecting the 80% of the annotated DT rows for the training and save the remaining 20% for the test. The DT row selection has been made by keeping the proportions among the classes. Then, the SVM parameters have been estimated by employing the *k*-fold method with *k*=5.

Since the great amount of variables causes a decrease in classification performances and accuracy, the approach includes ways to reduce dimensions and improve performances. For this purpose, the Principal Component Analysis (PCA) has been used to extract linearly independent variables at maximum variance from the original dataset. The effective number of components to pick is determined through cross-validation. Data achieved through PCA-based reduction are also standardised to boost SVM accuracy. Once that cross-validation, PCA-based data reduction, and standardisation have been accomplished, the model is finally trained.

3.3.3 A "soft" classification

The human mood is generally a mixture of different emotions, with different intensity degrees. In many cases, a feeling is not easy to detect and label with a specific name, because it is composed of various sensations and sentiments that are not easy to interpret. Keeping this in mind, the proposed approach performs a soft classification of the hashtags to improve detection and description of people's emotions and reactions to events, that in this case, are represented by the relative hashtag trends. The proposed classification approach employs multi-class SVM that is based on pairwise classification, which associates a classifier to each class pair. Therefore, the classification in m classes is performed by using $m \cdot (m-1)$ classifiers. In this scheme, each classifier picks one class over the other from a pair of classes. Counting the number of times a class is preferred on all the pairs, the algorithm assesses the number of votes that the class received. In the traditional hard classification approaches, the class that reaches the highest number of votes is selected and associated with the hashtag (majority voting algorithm). Our soft classification approach, instead, not only considers the most voted class to classify a hashtag, but also takes into account the votes generated for each class over the $m \cdot (m-1)$ pairs. The number of votes generated for the ETL emotional class E is normalised in the range $[0,1]$ and expresses the intensity of the feeling represented by E. The set of all the votes for each ETL emotional class is the output of the hashtag soft classification, which represents the wide and complex emotional reactions expressed by people towards something (i.e., hashtag associated with an event, topic, etc.). As an example of classification for a tweet trend (i.e., hashtag), let us consider the following: ((open 0.0), (happy 0.3), (alive, 0.0), (good 0.4), (love 0.2), (interest 0.0), (positive 0.0), (strong 0.0), (angry 0.1), (depressed 0.0), (confused 0.0), (helpless 0.0), (indifferent 0.0), (afraid 0.0), (hurt 0.0), (sad 0.0).

This set shows the ETL emotional classes, each one of them annotated with the intensity degree determined through classification, that depict the emotional reactions to a tweet trend identified by a specific hashtag. According to the values, it appears that the comprehensive feeling towards the hashtag considered is pleasant due to the high intensities achieved for some emotional classes, such as *good* and *happy*, from the analysis of tweets about that hashtag. Other classes, such as *love*, contribute to describing the nuances of the human mood fully. The soft classification returns a full description of all the feelings expressed by people about an event, contrary to the hard classification that only detects the dominant feeling (i.e., in the example above, it would have classified the reaction as good, even though there is a minority of angry people). The soft classification provides a comprehensive description of all the emotional reactions expressed by people towards a hashtag/event.

3.4 Ontology-based emotion view

The *Ontology-based emotion view* module in the architecture (see Figure 1) provides semantic annotations on all the extracted emotional concepts.

An ontology of the emotional concepts, namely the *Emotional Concept Ontology*, is built by generating assertions over the BCs and CETs, extracted through the simplicial complex structure. The knowledge base built over the emotional concepts can enrich the classification results on the emotional reactions. Meaningful insights on the emotional concepts can be provided by querying the generated knowledge base. These insights can better depict the sentiments and reactions towards the hashtags/events taken into consideration.

3.4.1 DBpedia-driven concept extension

Our ontology concepts are enriched by adding semantic relations with DBpedia,[2] an online semantic dataset for Wikipedia structured content. In other words, our approach gets additional relations by DBpedia and enrich our ontology concepts corresponding to BCs and CETs with related well-defined DBpedia concepts. SPARQL2[3] queries are defined to search for DBpedia concepts related to the BCs. Emotional concepts extracted from DBpedia are related to each BC in our ontology model. To clarify this step, let us consider the SPARQL query example in Figure 5; the DBpedia emotional concepts related to the BC containing the (stemmed) term "happi" are retrieved. Each resource subcategory of *dbc:Emotion* (line 4) with a label (line 5) containing "happi" (line 7) is explored.

Figure 5 SPARQL query to extract DBpedia concepts related to the BC "happi"

```
1.  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2.  SELECT DISTINCT ?res
3.  WHERE{
4.  ?res dct:subject dbc:Emotions.
5.  ?res rdfs:label ?lbl
6.  FILTER(lang(?lbl)='en')
7.  FILTER(regex(str(?lbl),'happi', "i")) }
```

res
http://dbpedia.org/resource/Happiness
http://dbpedia.org/resource/Aversion_to_happiness

3.4.2 Emotional concept ontology in SKOS

The extracted BCs and CETs, enriched by DBpedia links, are encoded into ontological resources to build the emotional concept ontology. The proposed model employs the Simple Knowledge Organisation System (SKOS)[4] model, which gives specifications to support the use of Knowledge Organisation Systems (i.e., thesauri, classification schemes, and taxonomies). SKOS provides a knowledge model to represent high-level concepts and the semantic relations among them. In our definition of the Emotional Concept Ontology, four properties have been taken into consideration to describe the emotional concepts:

- *skos:related*: a SKOS property representing a broad basic relationship among concepts
- *skos:narrower*: a SKOS property used to express a closer relationship among concepts

- *rdfs:seeAlso*: a RDF property to add additional information to a RDF resource.
- *skos:member*: a SKOS property to relate a SKOS concept to a SKOS Collection.

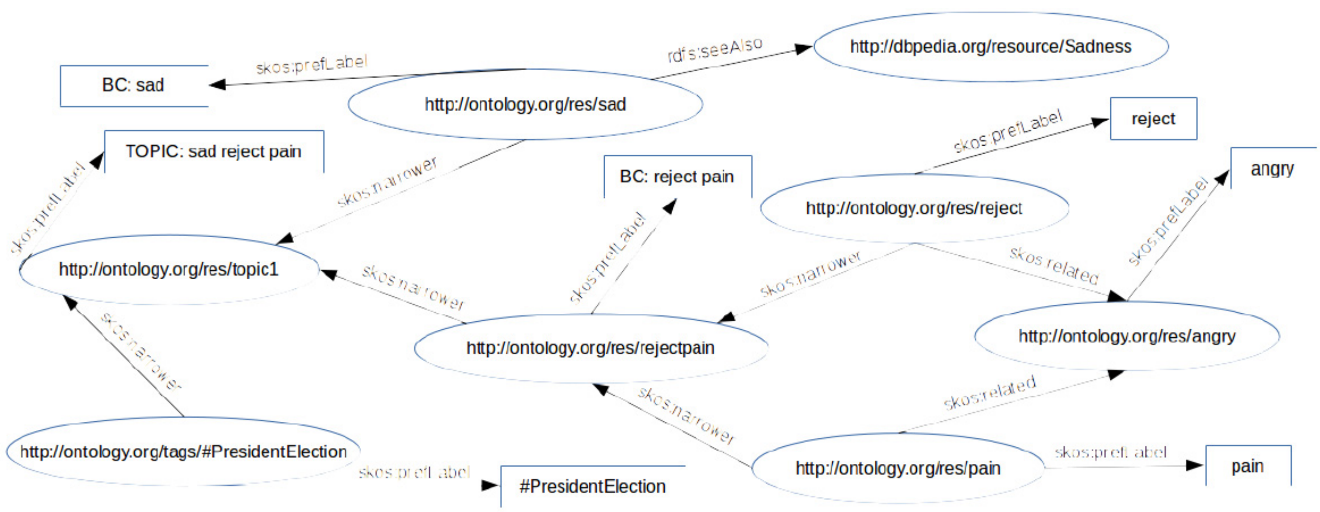
In this ontology modelling, the class *skos:Concept* describes the emotion as sets of terms (i.e., the emotion is a blend of all the term-based emotional nuances) automatically extracted by the framework. In detail, the hashtags, the topics, the BCs and CETs extracted via simplicial complex and the DBpedia concepts matching the found BCs are all encoded as instances of the class *skos:Concept*. Each emotional class (e.g., HAPPY, ANGRY) in ETL, also used for the hashtag emotional classification, is modelled as an instance of the class *skos:Collection* representing the vocabulary of terms the emotional class is composed of. All these SKOS-based concepts allow building a concept hierarchy based on relations among the BCs, and extend each BC with the CETs and the DBpedia matching resources (see Sub-section 3.4.1). The hashtags are related to the topics they have been associated with. The topic is in turn, described by the BCs and their extended related knowledge (i.e., CETs and DBpedia concepts).

To this purpose, the property *skos:narrower* links, as usual, all those concepts that are strictly related one another by a hierarchical relationship. In detail, the property is used to represent the relation between the hashtag resource and its associated emotional topic, composed of all the BCs in the hashtag document. In this case, the property *skos:narrower* represents the strict and deep relationship between the hashtag and the emotional topic. The *skos:narrower* also describes the membership of a BC to a topic, and also relates each term to the BC which is part of. The property *rdfs:seeAlso* is used to relate each BC (*skos:Concept*) to the DBpedia resource expressing a related concept (*skos:Concept*). The DBpedia resource generally adds much more information about the BC (i.e., definitions, related concepts, and BC contexts).

Each BC is also related to the CETs. As stated, CETs are roughly related to the BC and provide a contextualisation of the BC (see Sub-section 3.2). Then, BC terms are linked to the CETs through the property *skos:related*. This way, all the resources linked to a BC term through *skos:related* are resources providing contextualised information about the BC. Each term, modelled as a *skos:Concept*, is related to the emotional class it belongs to. Recalling that the emotional classes used for the classification are modelled as vocabularies of terms through the class *skos:Collection*, terms, related to a BC via the property *skos:narrower*, are also related to the vocabulary/emotional class to which they belong through the property *skos:member*.

The introduced ontology model provides the basic knowledge representation to relate emotional concepts, reactions and events. Data achieved through the simplicial complex structure and the classification are used to populate the ontology with facts about the BCs, CETs, tweet trends and the extracted emotional valences. An extract of the emotional concept ontology is shown in Figure 6, it represents emotional concepts extracted from documents related to the trend with the hashtag *#PresidentElection*. The hashtag is associated with its emotional topic *topic1*, through the property *skos:narrower*; by using the same property, the BCs *sad* and *rejectpain*, which are present in *topic1*, are related to this topic. The *reject* and *pain* terms are related, in turn, to the BC *rejectpain* through the *skos:narrower*. Related DBpedia resources, if detected (see Sub-section 3.4.1), are also related to the BCs via the property *rdfs:seeAlso*; for instance, the BC *sad* is related to the DBpedia concept *Sadness*. If CETs extending the considered BCs are present, they are related to them through the property *skos:related*, such as the BC terms *reject* and *pain* that are associated to the common CET *angri*. The description of each ontology resource is given by the SKOS label.

Figure 6 Emotional concept ontology extract



3.4.3 Query-generated insights on emotional reactions

Once the knowledge base is built, views on the model can be performed with SPARQL queries. These views can provide insights on the hashtag emotion classification to better support the analysis of the emotions and reactions towards the event/hashtag. Queries can generate various views according to the application and type of analysis taken into consideration. In most of the cases, the interpretation of the emotional reactions towards an event is non-trivial and requires ways to define a context on the reactions. Through classification, the approach detects the most dominant reactions, but does not support the explanation and interpretation of the detected reactions. Queries on the built knowledge base allow to serve reaction explanation by retrieving all the emotional concepts (BCs) related to the event or product (hashtag) and, then, build the contextually-related terms to better support the interpretation of the emotional reactions. In details, the query, shown in Listing 1, allows the detection of the topic composed of all the emotional BC terms (line 9) related to the event Billboard Music Awards (hashtag #BBMAs, line 7). Then, the query in Listing 1 returns the CETs associated to each BC term (line 9) in the topic associated to the hashtag #BBMAs. In a similar way, the query in Listing 1 builds and returns the list of the DBpedia concepts related to the BC terms (line 9) present in the topic associated to the hashtag #BBMAs. The so-extracted lists of CETs and DBpedia concepts define a context on the terms of the topic associated to the hashtag. This term-set-based context provides a way to support the interpretation of reactions to the event.

Other kinds of queries can support other applications, such as the individuation of which events cause the same people's reactions, in order to support the analysis of the sociological context or the impact generated by the launch of a new product on potential customers. To this purpose, the query in the Listing 2 returns the list of all the hashtags, and their own topics (line 7), to which people reacted with happiness (lines 8 and 9).

Some views can also support the detection of the events to which people's reactions are very controversial. To support this application, the query in Listing 5 counts all the terms in a topic, and returns a list of the hashtags ordered according to the number of terms present in their topics. Therefore, the last hashtags in the list (i.e., those with the greatest number of terms in their topics) are those events to which people react in the most different way. Therefore, those events may be easily selected to be better analysed.

There are many other kinds of query that can improve classification results and support the hashtag emotional analysis. Thanks to the simplcial complex structure and the connection with DBpedia, the ontology model built on these data allows to serve many applications.

4 Approach evaluation

This section shows possible benefits of using the soft classification along with the ontological model to detect people's reactions to events from tweets.

4.1 Data set and system configuration

Our approach has been tested on tweet streams containing hashtags in Twitter top trends, collected from June to August 2018. Tweets are related to hashtags that refer to several events, including generic ones, such as weekly thoughts and memories (e.g. *#WednesdayWisdom*, *#TBT*, *#ThursdayThoughts*, *#FridayFeeling*) and more specific events, from important international political events (e.g., *#PMQs*, *#PrimaryElection*) to lighter events (e.g., *#NationalChoccolateIceCreamDay*, *#NationalBestFriendDay*). The most common hashtag topics include political, social and religious news and events (e.g., *#politicalcorrectness*, *#RSSTritiyaVarsh*), opinions on famous people and politicians (e.g., *#Trump*), campaigns and initiatives (e.g., *#FlirtWithYourCity*, *#GlobalRunningDay*), films and TV series (e.g. *#astarisborn*, *#loveisland*) and important events in music and fashion (e.g., *#TonyAwards*, *#WMA*, *#GNTM*).

The collected tweets have been arranged in documents, each one dedicated to a hashtag, and stored. The tests have been processed on a computer equipped with 32 GB ram, 2 TB SATA HDD and 8 quadcores, to support the acquisition of high amounts of tweets and speed up the execution of the whole system. The hashtag document corpus is composed of 5617 documents, each one of them made up of tweets about a specific hashtag. The tweets, present in the whole corpus, are around 11,234,000, with 2000 tweets per document, on average.

4.2 Classification evaluation

Classification test results have been compared against the simplicial complex output. In detail, the structure returns the topic T (i.e., set of emotional BCs and terms) that is associated with each tweet document D in the test set. Then, each document D has been annotated with the class in ETL, including the greatest number of terms in its topic T. This solution has been proven to be effective for evaluating the efficiency of sentiment analysis methods (Di Martino et al., 2019).

Table 1 Experimental results by varying configuration setting

Exp	*Doc*	*Feats*	*plF*	*unF*	*plC*	*unC*	*F1*
1	5617	250	126	124	7	6	0.55
2	5617	100	38	37	6	5	0.57
3	5617	250	128	122	6	5	0.62
4	5061	100	100	0	6	0	**0.64**
5	5061	109	58	51	6	6	0.55
6	5061	109	58	51	6	5	0.58
7	5061	100	58	51	6	4	0.59
8	5061	100	58	51	6	0	0.67
9	4000	105	54	51	6	5	**0.7**
10	4000	70	36	34	5	4	**0.7**
11	4000	46	16	30	2	3	**0.81**
12	4000	30	11	19	6	2	**0.89**
13	4000	30	11	19	2	3	**0.97**
14	3000	150	150	0	6	0	**0.68**
15	2000	30	16	14	6	4	**0.90**
16	2000	150	64	86	6	7	0.64

Table 2 Experimental results by varying configuration setting

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	**0.70**
5	5061	109	58	51	6	6	0.55
6	4000	105	54	51	6	5	**0.7**
7	4000	70	36	34	5	4	**0.7**
8	4000	46	16	30	2	3	**0.81**
9	4000	30	11	19	6	2	**0.89**
10	4000	30	11	19	2	3	**0.97**
11	3000	150	150	0	6	0	**0.68**
12	2000	30	16	14	6	4	**0.90**
13	2000	150	64	86	6	7	0.55

Classification results are shown in Table 3, where each row represents an experiment (Exp), described by the number of processed documents (doc), the number of features (feats), and also split into pleasant (plF) and unpleasant (unF) ones. Then, the number of pleasant (plC) and unpleasant (unC) classes considered for the experiments are also given. F1 score (F1) has been reported for each experiment (Exp). Classes with small numbers of features have been removed, as they are considered irrelevant to the whole corpus. Experiments conducted on pleasant features exclusively, such as *Exp 4*, return better F1 score values than experiments carried out on both the pleasant and unpleasant features, such as *Exp 5*, even though both the types of experiment are done over the same number of documents and almost the same number of features. This result can be explained by the fact that pleasant features appear more frequently in tweet documents than unpleasant ones, mainly depending on the tweet trend content, to which people tend to reply with positive assertions.

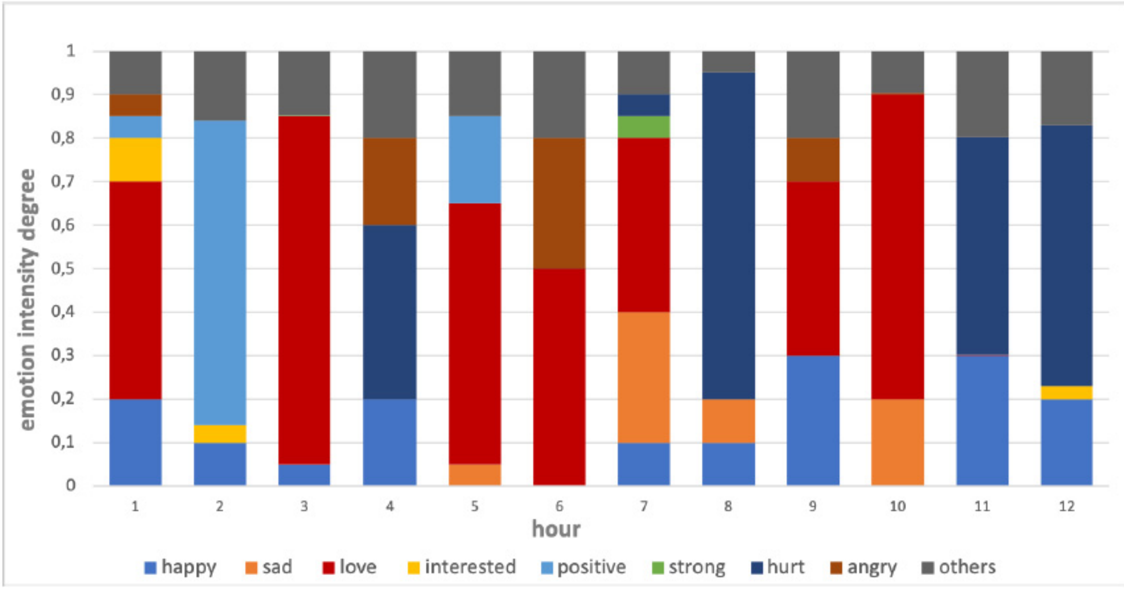
Table 3 Experimental results by varying configuration setting

Exp	*Doc*	*Feats*	*plF*	*unF*	*plC*	*unC*	*F1*
1	5617	250	126	124	7	6	0.55
2	5617	75	38	37	6	5	0.57
3	5617	250	128	122	6	5	0.60
4	5061	100	100	0	6	0	**0.70**
5	5061	109	58	51	6	6	0.55
6	4000	105	54	51	6	5	**0.70**
7	4000	70	36	34	5	4	**0.70**
8	4000	46	16	30	2	3	**0.80**
9	4000	30	11	19	6	2	**0.87**
10	4000	30	11	19	2	3	**0.95**
11	3000	150	150	0	6	0	**0.67**
12	2000	30	16	14	6	4	**0.90**
13	2000	150	64	86	6	7	0.64

The F1 score is quite lower in some experiments, such as in *Exp 13*: analysing the data, it turns out that the terms such as *open* and *alive*, related to pleasant classes, do not refer to emotions but are very generic adjectives. Consequently, removing these terms generates a considerable improvement in the classification results (see *Exp 12*). To this purpose, PCA-based feature reduction has been carried out on most of the experiments reported in the table. Experiments *Exp 8*, *Exp 9* and *Exp 10*, for instance, show better performance in terms of F1 score, compared with the F1 score of the experiments *Exp 1*, *Exp 2* and *Exp 3* (where no feature reduction has been applied). Let us notice that, by keeping the document size fixed, the feature reduction applied to the experiments from *Exp 6* to *Exp 9* causes a significant increase in the F1 score. The selected features keep classification with high F1 score values, even though the number of documents changes. Let us notice this trend in experiments *Exp 12* and *Exp 9*, where the features are around 30 and the F1 score is stable around 90% in both the experiments, even though the number of documents grows from 2000 of *Exp 12* to 4000 of *Exp 9*. The emotional classification of the hashtags allows analysing people's emotional reactions to events over time. Figure 7 shows a chart bar graph describing people's reactions to World Cup 2018, as emotional states expressed in tweets with the hashtag #worldcup2018. The tweets have been retrieved from the streaming by starting from the first tweet about the event, and grouping tweets on the events per hour. Each bar depicts people's reaction at a specific hour from the start of the event. The bar sequence describes the evolution of people's reactions over time. Let us notice from these results that the two sentiments, namely *Love* and *Hurt* alternate as time goes by. This trend describes supporters' emotional reactions to the performances of their teams in matches, i.e., supporters reacted by expressing positive feelings if their team won (i.e., *Love*), negative feelings otherwise (i.e., *Hurt*).

The different widths of the colored areas composing each bar of Figure 7 describe the soft classification of the emotional reactions to the hashtag at each hour. In detail, the largest colored area in the bar at the hour h represents the highest-ranked emotional class, i.e., the most dominant emotional reaction to the event in that hour. The remaining smaller areas in the same bar, describe all the other emotions. This soft classification allows detecting different trends in reactions to events. For instance, at hour 6, even though the dominant emotional reaction is described by a positive feeling (i.e., *Love*), the second highest-ranked reaction is a negative feeling (i.e., *Angry*). The two major reactions, i.e., the two highest-ranked sentiments, have often close ranking values. This trend happens quite often, such as for *Love* and *Angry* at hour 6, *Love* and *Interested* at hour 7, *Love* and *Happy* at hour 9, etc.

Figure 7 Emotions associated with the hashtag #worldcup2018 at each hour from the first appearance of the hashtag



4.3 Insights on reactions through ontology-based views

Once the classification is carried out, the *Emotional Concept Ontology* is populated with the extracted emotional concepts (BCs and CETs) and DBpedia data. This knowledge base can be used to generate views on reactions through SPARQL queries. These views can provide insights into the classification to better support the analysis of the emotions and reactions towards the event/hashtag.

In most cases, the interpretation of the emotional reactions towards an event is non-trivial and requires ways to define a context on the reactions. Through classification, the approach detects the most dominant reactions, even though it does not support the explanation and interpretation of the reactions. For instance, let us consider the classification results achieved by classification on the hashtag #worldcup2018 at hour 2 (see Figure 7): ((open 0.023), (happy 0.1), (alive, 0.022), (good 0.025), (love 0.02), (interest 0.03), (positive 0.7), (strong 0.01), (angry 0.01), (depressed 0.0), (confused 0.028), (helpless 0.0), (indifferent 0.0), (afraid 0.022), (hurt 0.0), (sad 0.01). The classification results evidences that the people's reactions towards #worldcup2018 at the hour 2 are mainly positive (e.g., positive, happy) with a minority of *afraid* and *confused* emotions, that are not easy to interpret. Queries on the knowledge base can support the reaction explanation by retrieving all the emotional concepts (BCs) related to the event (hashtag) and, then, build lists of CETs and DBpedia concepts, related to the BCs, to better support the interpretation of the emotional reactions. The query in Listing 1 retrieves the topic composed of all the emotional BC terms (line 9) related to the event at hour 2 for the hashtag #worldcup2018 (line 7 , i.e., all the tweets with the hashtag #worldcup2018_2 at hour 2 are stored to be easily retrieved). Then, the BCs in the topic along with their emotional classes of membership are retrieved as shown in Table 4. BCs are specific emotional terms depicting the emotional reaction classes determined through classification. For instance, beyond the term "good", the terms "calm" and "relax" contribute to better explain the emotional reaction expressed with the emotional class *Good*.

Listing 1: Query to detect the BC terms in the topic referring to the hashtag #worldcup2018, and their emotional class.

```
1  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
3
4  SELECT DISTINCT ?classLabl ?BClabl
5
6  WHERE {
7       ?hashtag skos:prefLabel '#worldcup2018_2'.
8       ?hashtag skos:narrower ?topic .
9       ?BCterm skos:narrower ?topic .
10      ?BCterm skos:prefLabel ?BClabl .
11      ?BCterm skos:member ?emClass .
12      ?emClass skos:prefLabel ?classLabl .
13 }
```

Table 4 Results of query in Listing 1: BCs in the topic of the hashtag #worldcup2018

Class	*BC*
Afraid	'http://ontology.org/res/terrifi'
Good	'http://ontology.org/res/relax'
Open	'http://ontology.org/res/free'
Confused	'http://ontology.org/res/embarrass'
Good	'http://ontology.org/res/good'
Good	'http://ontology.org/res/calm'
Positive	'http://ontology.org/res/optimist'
Strong	'http://ontology.org/res/strong'
Open	'http://ontology.org/res/confid'
Alive	'http://ontology.org/res/play'
Positive	'http://ontology.org/res/excit'
Happy	'http://ontology.org/res/great'
Afraid	'http://ontology.org/res/nervous'
Alive	'http://ontology.org/res/thrill'
Happy	'http://ontology.org/res/cheer'
Positive	'http://ontology.org/res/dare'
Alive	'http://ontology.org/res/anim'
Love	'http://ontology.org/res/love'
Happy	'http://ontology.org/res/jubil'
Interested	'http://ontology.org/res/intrigu'
Positive	'http://ontology.org/res/keen'
Interested	'http://ontology.org/res/interest'
Positive	'http://ontology.org/res/enthusiast'
Happy	'http://ontology.org/res/lucki'
Alive	'http://ontology.org/res/aliv'
Confused	'http://ontology.org/res/tens'
Good	'http://ontology.org/res/clever'

To improve reaction explanation, queries to detect CETs associated with BCs are shown as well. The query in Listing 2 returns the CETs related to each BC term (line 12) in the topic (line 9) associated with the hashtag #worldcup2018 (lines 6 and 7).

Table 5 reports the list of the CETs returned by the query in Listing 2. Let us notice that the CETs in Table 5 associated with each BC term better explain the reactions. For instance, the CETs "dismay", "incens" and "agon" in Table 5 refine the BC "nervous". Queries can also discover DBpedia concepts related to the detected BCs (see Appendix A), to enrich the knowledge generated on people's emotional reactions with concepts from the DBpedia graph.

Listing 2: Query to detect the CETs related to BCs in the topic associated to the hashtag #worldcup2018.

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-
      schema#>
2 PREFIX skos: <http://www.w3.org/2004/02/skos/
      core#>
3
4 SELECT DISTINCT ?term ?CET
5 WHERE {
6     ?hashtag skos:prefLabel '#worldcup2018_2'.
7     ?hashtag skos:narrower ?topic .
8     ?topic skos:prefLabel ?lablTP .
9     ?BCterm skos:narrower ?topic .
10    ?BCterm skos:prefLabel ?lablBC .
11    ?term skos:narrower ?BCterm .
12    ?term skos:related ?CET .
13 }
```

Table 5 Results of query in Listing 2: CETs related to five BCs (embarass, cheer, clever, tens, nervous) in the topic of the hashtag #worldcup2018

BC	*CET*
3*'http://ontology.org/res/embarass'	'http://ontology.org/res/perplex'
	'http://ontology.org/res/disillus'
	'http://ontology.org/res/stupefi'
2*'http://ontology.org/res/cheer'	'http://ontology.org/res/provoc'
	'http://ontology.org/res/elat'
3*'http://ontology.org/res/clever'	'http://ontology.org/res/tenaci'
	'http://ontology.org/res/rebelli'
	'http://ontology.org/res/engross'
4*'http://ontology.org/res/tens'	'http://ontology.org/res/indign'
	'http://ontology.org/res/detest'
	'http://ontology.org/res/infuri'
	'http://ontology.org/res/restless'
3*'http://ontology.org/res/nervous'	'http://ontology.org/res/dismay'
	'http://ontology.org/res/incens'
	'http://ontology.org/res/agon'

Further queries can be formulated to support the analysis of emotional reactions to events/hashtags; and thanks to the conceptual connection with DBpedia, the ontology model is suitable for several purposes (see Appendix A for more details).

5 Conclusions

The paper introduced a hybrid approach to extract people's emotional reactions towards events from tweets by using emotional classification and an emotional concept ontology. The approach builds a linguistic topological space of the emotional terms extracted from tweets by using the geometrical structure simplicial complex. Through the structure, emotional basic concepts (BCs) are extracted and used to train a Support Vector Machine (SVM) classifier aimed at detecting emotional classes describing people's reactions towards hashtags/events.

The ontology is built and populated with the BCs and their CETs, extracted by the simplicial complex, whereas the emotional reactions are detected by classification results. The ontology can be queried to generate meaningful insights on emotions by integrating BCs, CETs, and DBpedia concepts matching the found BCs, to support the analysis and explanation of people's reactions.

The paper contribution is manifold. Firstly, the proposed approach allows emotional reaction analysis by combining Machine Learning methods with Semantic Web technologies to detect the emotional reactions from tweets and support their analysis by querying the knowledge base built on the detected people's reactions.

Furthermore, the framework does not only individuate and describe the main reactions but also provides solutions to accomplish complex reaction analyses. From the tweet analysis, indeed, the knowledge base can be used to analyse the retrospective impacts of the event and, accordingly, support various applications, including voting intention analysis, social movement spreading, and marketing (i.e., reaction evaluation of new product release on the market).

Tests carried out on the analysis of reactions have shown the benefits of the soft classification of the emotions. The soft classification has been proven to achieve a more significant and detailed description of people's emotional reactions towards the events. Furthermore, the reaction classification may be further investigated by submitting simple queries over the knowledge base: meaningful views on reactions can be generated to support human experts in their event impact analysis.

Many open challenges need to be addressed to achieve thorough emotional reaction analysis, especially the detection and disambiguation of mixed emotions. Mixed emotions are those emotions that cannot be exclusively classified as positive, negative, or neutral sentiments since they represent emotions in between these categories. Additionally, the presence of irony in tweets may also complicate the detection of emotional reactions. To this purpose, future works will focus on extensions of the proposed framework for emotional reaction detection to take into account the analysis of mixed emotions and ironic expressions.

References

Alharbi, A.S.M. and De Doncker, E. (2019) 'Twitter sentiment analysis with a deep neural network: an enhanced approach using user behavioral information', *Cognitive Systems Research*, Vol. 54, pp.50–61.

AL-Sharuee, M.T., Liu, F. and Pratama, M. (2018) 'Sentiment analysis: an automatic contextual analysis and ensemble clustering approach and comparison', *Data and Knowledge Engineering*, Vol. 115, pp.194–213.

Basha, S.M. and Rajput, D.S. (2018) 'A supervised aspect level sentiment model to predict overall sentiment on tweeter documents', *International Journal of Metadata, Semantics and Ontologies*, Vol. 13, No. 1, pp.33–41.

Çal, S. and Balaman, S.Y. (2019) 'Improved decisions for marketing, supply and purchasing: mining big data through an integration of sentiment analysis and intuitionistic fuzzy multi criteria assessment', *Computers and Industrial Engineering*, Vol. 129, pp.315–332.

Cavaliere, D. and Senatore, S. (2019) 'Emotional concept extraction through ontology-enhanced classification', in Garoufallou, E., Fallucchi, F. and De Luca, E.W. (Eds): *Proceedings of the 13th International Conference Metadata and Semantic Research (MTSR'19)*, Rome, Italy.

Cavaliere, D., Senatore, S. and Loia, V. (2017) 'Context-aware profiling of concepts from a semantic topological space', *Knowledge-Based Systems*, Vol. 130, pp.102–115.

Cotelo, J.M., Cruz, F.L., EnrÃ-quez, F. and Troyano, J.A. (2016) 'Tweet categorization by combining content and structural knowledge', *Information Fusion*, Vol. 31, pp.54–64.

Di Martino, F., Senatore, S. and Sessa, S. (2019) 'A lightweight clustering-based approach to discover different emotional shades from social message streams', *International Journal of Intelligent Systems*, Vol. 34, No. 7, pp.1505–1523.

Hussein, D.M.E.D.M. (2018) 'A survey on sentiment analysis challenges', *Journal of King Saud University – Engineering Sciences*, Vol. 30, No. 4, pp.330–338.

Ibtihel, B.L., Lobna, H. and Maher, B.J. (2018) 'A semantic approach for tweet categorization', *Procedia Computer Science*, Vol. 126, pp.335–344.

Li, X., Feng, S., Wang, D. and Zhang, Y. (2019b) 'Context-aware emotion cause analysis with multi-attention-based neural network', *Knowledge-Based Systems*, 174, pp.205–218.

Li, X., Wu, C. and Mai, F. (2019a) 'The effect of online reviews on product sales: a joint sentiment-topic analysis', *Information & Management*, Vol. 56, No. 2, pp.172–184.

Öztürk, N. and Ayvaz, S. (2018) 'Sentiment analysis on twitter: a text mining approach to the Syrian refugee crisis', *Telematics and Informatics*, Vol. 35, No. 1, pp.136–147.

Poria, S., Cambria, E., Winterstein, G. and Huang, G.B. (2014) 'Sentic patterns: dependency-based rules for concept-level sentiment analysis', *Knowledge-Based Systems*, Vol. 69, pp.45–63.

Santosh, D.T. and Vardhan, B.V. (2016) 'PROO ontology development for learning feature specific sentiment relationship rules on reviews categorisation: a semantic data mining approach', *International Journal of Metadata, Semantics and Ontologies*, Vol. 11, No. 1, pp.29–38.

Singh, F. and Kumari, M. (2016) 'Role of text pre-processing in twitter sentiment analysis', *Procedia Computer Science*, Vol. 89, pp.549–554.

Symeonidis, S., Effrosynidis, D. and Arampatzis, A. (2018) 'A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis', *Expert Systems with Applications*, Vol. 110, pp.298–310.

Thakor, P. and Sasi, S. (2015) 'Ontology-based sentiment analysis process for social media content', *Procedia Computer Science*, Vol. 53, pp.199–207.

Yaqub, U., Chun, S.A., Atluri, V. and Vaidya, J. (2017) 'Analysis of political discourse on twitter in the context of the 2016 US presidential elections', *Government Information Quarterly*, Vol. 34, No. 4, pp.613–626.

Yoo, S.Y., Song, J. and Jeong, O.R. (2018) 'Social media contents based sentiment analysis and prediction system', *Expert Systems with Applications*, Vol. 105, pp.102–111.

Zhang, F., Fleyeh, H., Wang, X. and Lu, M. (2019) 'Construction site accident analysis using text mining and natural language processing techniques', *Automation in Construction*, Vol. 99, pp.238–248.

Notes

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

2 https://wiki.dbpedia.org/

3 https://www.w3.org/TR/sparql11-query/

4 https://www.w3.org/2004/02/skos/

Appendix

A. Queries and applications

To enrich the knowledge based on BCs and CETs, DBpedia concepts can be retrieved through queries. To this purpose, the query in Listing 3 returns the list of the DBpedia concepts related to the BC terms (line 11) present in the topic (line 9) associated to the hashtag #worldcup2018 (line 7). An example of query output is present in Table 6, showing a list of all the BCs that match feelings and effective emotional concepts on DBpedia, linking all the knowledge built on the emotional reactions by the approach to the DBpedia graph.

Listing 3: Query to detect the DBpedia concepts matching with the BCs in the topic associated to the hashtag #worldcup2018

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
3
4 SELECT DISTINCT ?lablBC ?DBpediaC
5 WHERE {
6     ?hashtag skos:prefLabel '#worldcup2018_2'.
7     ?hashtag skos:narrower ?topic .
8     ?topic skos:prefLabel ?lablTP .
9     ?BCterm skos:narrower ?topic .
10    ?BCterm skos:prefLabel ?lablBC .
11    ?BCterm rdfs:seeAlso ?DBpediaC .
12 }
```

Table 6 Results of query in Listing 3: some DBpedia concepts related to the BCs in the topic of the hashtag #worldcup2018

BC	*DBpedia concept*
2*'http://ontology.org/res/passion'	'http://dbpedia.org/resource/Compassion'
	'http://dbpedia.org/resource/Passion_(emotion)'
'http://ontology.org/res/surpris'	'http://dbpedia.org/resource/Surprise_(emotion)'
2*'http://ontology.org/res/sure'	'http://dbpedia.org/resource/Measures_of_guilt_and_shame'
	'http://dbpedia.org/resource/Pleasure'
3*'http://ontology.org/res/eas'	'http://dbpedia.org/resource/Pleasure'
	'http://dbpedia.org/resource/Measures_of_guilt_and_shame'
	'http://dbpedia.org/resource/Reasonable_person_model'
'http://ontology.org/res/interest'	'http://dbpedia.org/resource/Interest_(emotion)'
3*'http://ontology.org/res/love'	'http://dbpedia.org/resource/Love'
	'http://dbpedia.org/resource/Romance_(love)'
	'http://dbpedia.org/resource/Love_at_first_sight'

Other kinds of queries can support other applications, such as the individuation of which events cause the same people's reactions, in order to support the analysis of the sociological context related to recent events or the impact generated by the launch of a new product on potential customers. To this purpose, the query in the Listing 4 returns the list of all the hashtags, and their own topics (line 7), to which people reacted with calmness (lines 8 and 9). Let us notice from query results in Table 7, that people react more often with calmness to special occasions, such as celebration days (e.g., #LoveYourPetDay, #UKPunDay, #NationalPizzaDay).

Listing 4: Query to detect the hashtag to which people react with calmness

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
3
4 SELECT DISTINCT ?lablH ?topic
5 WHERE {
6     ?hashtag skos:prefLabel ?lablH .
7     ?hashtag skos:narrower ?topic .
8     ?term skos:narrower ?topic.
9     FILTER(regex(str(?term),'calm', "i")).
10 }
```

Table 7 Results of query in Listing 4: hashtags to which people reacted with calmness

Hashtag
'#NSD18'
'#SingleValentinesPlans'
'#LoveYourPetDay'
'#FalconHeavy'
'#UKPunDay'
'#AJL32'
'#WomenInScience'
'#SaturdayMorning'
'#NationalSigningDay'
'#GetMeHotIn4Words'
'#BlackPantherTheAlbum'
'#ARMYSelcaDay'
'#MondayMotivation'
'#StrayKidsStanSelcaDay'
'#SundayMorning'
'#TheBachelor'
'#SCPFCB'
'#Venom'
'#EaglesParade'
'#NZvENG'
'#NEWMUN'
'#NationalWeatherpersonsDay'
'#PutSomeRandomGuyInAFilm'
'#NationalPizzaDay'
'#biathlon'
'#InternationalClashDay'
'#Adam_Lallana'

Some views can also support the detection of the events to which people's reactions are very complex or even controversial in a specific period. To support this application, the query in Listing 5 counts all the terms in a topic, and returns a list of the hashtags ordered according to the number of terms present in their topics in descendent order. Therefore, the first hashtags in the list (i.e., those with the greatest number of terms in their topics) are those events to which people react in the most different way. Therefore, those events may be easily selected to be better analysed. For instance, as shown in the query results on Table 8, in tweets in the period going from June to August 2018, people react most differently when describe their own experiences about life and memories ("#TueasdayThoughts", "#TBT") or on facts on famous people (e.g., '#Quincy_Jones').

Listing 5: Query to return a list with hashtags ordered according to the number of their BCs

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT DISTINCT ?hashtag (count(?hashtag) as ?tagcount)
WHERE {
      ?hashtag skos:narrower ?topic .
      ?topic skos:prefLabel ?lablTP .
      ?term skos:narrower ?topic .
      ?term skos:prefLabel ?lablT .
      FILTER(regex(str(?lablTP),'TOPIC: ', "i"))
}
group by ?hashtag
order by ?tagcount
```

Table 8 Results of query in Listing 5: hashtags with the highest number of different reactions

Hashtag	*# reactions*
'#TuesdayThoughts'	'40'
'#MondayMotivation'	'36'
'#MondayMotivation'	'36'
'#ThursdayThoughts')	'36'
'#ModiHitsBack'	'33'
'#SundayMorning'	'31'
'#WednesdayWisdom'	'31'
'#TuesdayThoughts'	'29'
'#Moscow'	'29'
'#Quincy_Jones'	'29'
'#TBT'	'29'
'#MigunaDeported'	'28'
'#FalconHeavy'	'27'
'#100years'	'27'
'#Josh_McDaniels'	'27'