Automated Classification of Book Blurbs According to the Emotional Tags of the Social Network Zazie

Valentina Franzoni, Valentina Poggioni, and Fabiana Zollo

Department of Mathematics and Computer Science,
University of Perugia, Perugia, Italy
fabiana.zollo@gmail.com,
{valentina.franzoni, poggioni}@dmi.unipg.it

Abstract. Sentiment Analysis and Opinion Mining are receiving increasing attention in many sectors because knowing and predicting opinions of people is considered a strategic added value. In the last years an increasing attention has also been devoted to Emotion Recognition, often by developing automated systems that can associate user's emotions to texts, music or artworks. Zazie is an Italian social network for readers that introduces a new dimension on book characterization, the emotional icon tagging. Each book, besides user's comments and reviews, can be tagged with special icons, the MOODs, that are emotional tags chosen by the users. The aim of this work is to study the feasibility of an automated classification of books in Zazie according to the emotional tags, by means of the lexical analysis of book blurbs. A supervised learning approach is used to determine if a correlation between the characteristics of a book blurb and the emotional icons associated to the book by the users exists.

Keywords: sentiment analysis, emotion recognition, automated classification, machine learning

1 Introduction

In the last years an increasing attention has been addressed to Sentiment Analysis and Emotion Recognition, often by developing automated systems that can associate users' emotions to text, music or artwork and can interpret the subjective nature of emotional content. Several efforts have been devoted to music emotion classification and recognition whether from a research or a commercial point of view [1] [2] [3] [4]. The underlying idea in most of these works is to analyze the track, representing it by means of a number of features and applying machine learning algorithms to a given dataset to infer models. A critical point in these works is the choice of the emotional model used to represent emotions and moods [5]; the choices are generally split in continuous models (e.g., Valence and Arousal [6] [7]) or discrete models (e.g., Ekman categories [11]).

Moreover, in the wide scenario of text mining, several attempts have been made in order to associate emotions and moods to blogs [8], tales [9], newspapers titles [12] in several domains and contexts.

A particular and interesting domain that, to the best of our knowledge, has not already been investigated is the association of emotions to books. The idea of building a model for classifying books from an emotional point of view was born from *Zazie* [10], a social network for readers that, differently from other similar projects e.g., *aNobii* or *Goodreads*, introduces a new dimension for books description, the emotional icon tagging.

Starting from this context, it would be very innovative to provide Zazie users with an emotion-driven search within the social network. The necessity of such an automated system arises also from the presence of a lot of books that have not been tagged yet by the users, for which there is not any information besides the characteristics stored in the database i.e., title, author or publisher.

The first step of this research was focused on the selection of relevant attributes among those usable and available in Zazie to describe a book. We decided to analyze the book blurb because it can contain relevant emotional information. Since the blurb is generally written for attracting the reader, it can emphasize and highlight some book aspects and it can abound with emotional terms: it seems to be suitable for the automated recognition of emotions. On the other hand, a possible drawback is the introduction of a bias caused by the excessive use of words with a high emotional meaning, so the problem is not trivial. Moreover, the blurb represents an information always available on Zazie, regardless of user's opinions, reviews or tags. The main original contribution of this work is to determine if the book blurb reflects the same emotions that the reader can find in the book itself. The emotional model used for MOODs representation is directly provided by Zazie by means of its emotional tags (icons) and can be easily correlated to the well known discrete emotional models, such as the ones defined by Ekman [11] or Plutchik [17].

This paper is organized as follows: in the next section the social network Zazie and its model for emotional tagging are described; then, in Section 3 related works are presented. The features extraction phase and the dataset creation are presented in Sections 4 and 5, while the experimental results are described and discussed in Section 6. The paper ends with some conclusions and ideas for future developments and improvements.

2 Zazie

Social networks for books recently had a widespread diffusion e.g., *aNobii*, *Shelfari*, *Bookish*, *Goodreads* with the common aim of creating communities of readers, allowing them to share their opinions about books and to obtain suggestions and advices. Zazie is an Italian social network for books, which is constructed on the same model of aNobii, but differs from it and from the other social networks of book readers for the opportunity to share emotions through icon tags: in addition to classic information as the average vote, readers' reviews and comments, Zazie provides users with icons, called MOODs, which introduce a new emotional dimension.

Each book in Zazie can be tagged with two MOODs icons, selected in a set of 25 different climates related to the reader's opinion about the book or to the emotion induced by the book (7 examples are shown in Figure 1).

For the aim of this work only a subset of these 25 MOODs is taken into account: the attention has been devoted to the icons representing the emotions induced by the book: *angry*, *cry*, *love*, *sad*, *smile*, *think*.

Fig. 1: Examples of MOODs icons used by Zazie

3 Related Work

The diffusion of social media has contributed to generate a data flow that constitutes an important information container and that has determined an increasing interest in Sentiment Analysis, whose diffusion stems from the facility of fruition of these information and from their numerous applications e.g., for social behaviour studies, financial services, social and political events.

In 2013 a preliminary study on automated classification of books has been proposed by the authors in the master thesis [23], where Zazie is presented and the emotional analysis of the blurb is introduced.

In 2005 Gilad Mishne presented a study for the automated classification of blogs, basing on the moods tagged by the authors [8]. Starting from a huge collection of posts, Mishne demonstrated how the accuracy increased significantly with the increase of the quantity of data available for training and, although low, the results did not differ substantially from human performances in implementing the same task.

In the same year, Cecilia Ovesdotter Alm et al. [9] presented the *SNoW* architecture and explored the problem of the automated classification of 22 fairy tales of the Grimm brothers by means of Support Vector Machine (SVM) with respect to the basic emotions by Ekman [11]. The results of the experiments were encouraging.

In 2008 Rada Mihalcea and Carlo Strapparava [12] also used the Ekman emotions to present a series of experiments regarding the automatic analysis of emotions, contained in titles of newspapers. They described the construction of a large annotated dataset with respect to six basic emotions: anger, disgust, fear, joy, sadness and surprise. The authors proposed different methods of knowledge-based and corpus-based automated identification of these emotions in a text, trying to determine which were the best.

In 2012 Erik Cambria et al. [13] presented a project aimed to supply employees of the marketing environment for a new social media tool, allowing the management of semantic information and providing in this way the opportunity to capture the polarity of opinions and the emotional information associated with user-generated content. In particular, the authors decided to consider reviews related to mobile phones, because of the good quality and the large amount of comments available on the Web. Firstly in Cambria's work the comments have been analyzed using the *Sentic Computing* [29] multidisciplinary approach to Opinion Mining; then, the information has been codified for Sentiment Analysis on the basis of different ontologies; finally, the resulting knowledge base was made available for classification using an *ad hoc* website.

In the same year [15] Kirk Roberts et al. [15] introduced a corpus collected from Twitter with annotated micro-blog posts (i.e., tweets) annotated with seven emotions: anger, disgust, fear, joy, love, sadness, and surprise, analyzing how emotions are distributed in the data annotated and comparing it to the distributions in other emotion-annotated corpora. Moreover, they used the annotated corpus to train a classifier that automatically discovers the emotions in tweets and presented an analysis of the linguistic style used for expressing emotions.

In [14] Matteo Baldoni et al. presented *ArsEmotica*, an application software for associating the predominant emotions to artistic resources of a social tagging platform. The aim of the work was to extract a rich emotional semantics of tagged resources through an ontology driven approach, exploiting and combining available computational and sentiment lexicons with an ontology of emotional categories. In [18] Federico Bertola and Viviana Patti presented some achievements on the topic of social tagging related to artworks and museums within the ArsEmotica framework. Their focus was on eliciting sharable emotional meanings from visitors' tags in online collections, by interactively involving the users of the virtual communities in the process of capturing the latent emotions behind the tags, relying on methods and tools from a set of disciplines ranging from Semantic and Social Web to NLP. The aim was the creation of a semantic social space where artworks can be dynamically organized according to a new ontology of emotions inspired by the Plutchik's model of human emotions.

The Plutchik's model is also used in [17], where Jared Suttles and Nancy Ide present an experiment to identify emotions in tweets, classifying emotions according to a set of eight basic bipolar emotions defined by the Plutchik's wheel of emotions. This allowed to treat the multi-class problem of emotion classification as a binary problem for four opposing emotion pairs. They applied distant supervision, which has been shown to be an effective way to overcome the need for a large set of manually labeled data to produce accurate classifiers.

4 Preprocessing and Features Extraction

In addition to the features that can be naturally associated to a book, we focused on the book blurb i.e., the content in the back cover, in order to analyze it from an emotional point of view, aiming to extract a set of emotions that can represent the book itself. The choice of taking into account the blurb instead of analysing directly the whole content of the book derives from the obvious infeasibility of processing a so long text. Moreover, the entire content of the book could be not available, because of the dataset or for copyright issues. The idea to verify if the emotions extracted from the blurb can be relevant compared to the MOODs tagged from the users, is promising. However the blurb analysis is not immediate and different factors influence the choice of methods to be adopted.

Similarly to the case of newspapers or online newsletter titles [12], the blurb is written for attracting the reader and consequently it makes use of emotional terms, and seems to be suitable for the automated recognition of emotions. The blurb is particularly concise, unlike other kinds of texts, as blog posts, tales or articles; thereby it is not appropriate to think in terms of words frequency or in terms of measures which are directly correlated to the length of the text. Therefore MultiWordNet [19], an extension of WordNet [20] including information on Italian and English words, and in particular WordNet-Affect [21] have been used in order to extract a series of emotions starting from the blurb, taking advantage of the existing relations between WordNet synsets.

4.1 Blurb Analysis.

The blurb analysis has been realized in three main phases: preprocessing, extraction of emotions and reduction of emotions.

Preprocessing. In this phase a series of passages in order to normalize the text of the blurb were carried on, obtaining a suitable output for efficient processing. Firstly a *stop words deletion* was applied to all the terms of ordinary usage and not incisive within the analysis process e.g., articles and prepositions. Then, a phase of *tokenization* was carried on, extracting words ignoring punctuation marks and digits. Once drew all the words, it was necessary to reduce each inflected form to its canonical form, called *lemma*. This procedure, called *lemmatization*, has been realized by means of *Morph-it!* [22], a morphological resource for the Italian language. In Italian, in fact, there are some more linguistic issues to face than the English language. For instance, adjectives are declined in many ways, depending whether they refer to males or females, singular or plural: the lemmatization phase is basic to reduce the noise due to this variability. When a certain word can belong to more than one grammatical category, depending on its role within the sentence (e.g., noun or verb), all the related lemmata are kept. The output at the end of the preprocessing phase consists in a list of lemmata.

Extraction of Emotions. Once terminated the preprocessing phase, the extraction of emotions by means of WordNet-Affect was realized. Firstly it was necessary to retrieve for each lemma the WordNet synsets associated to it, using the multilinguale lexical databse MultiWordNet.

At this stage the affective domain WordNet-Affect was exploited in order to obtain all the emotions associated to the synsets, filtering out the terms which did not convey affective information and taking into account multiple occurrences of the same emotion.

Let us see an example: the blurb of *The Count of Montecristo*. The emotions extracted and their frequency are the following:
`negative-concern[1], horror[1], anxiety[1], distress[1], enthusiasm[1], negative-fear[1], love[1], affection[1], hate[1], comfortableness[1]`

Reduction of Emotions. It is necessary to highlight that the emotional hierarchy of WordNet-Affect is particularly pronged, with 296 nodes, and the result of the emotion extraction phase can be excessively detailed. For this reason the set of emotions has to be reduced and two different approaches have been implemented and tested.

At first the reduction has been made to the 32 emotions corresponding to the third level of hierarchy i.e., the subtree rooted in `emotion`. Each extracted emotion was replaced with the emotion reached by climbing the hierarchy up to the third level consisting of:

1. *love, affection, liking, enthusiasm, gratitude, self-pride, levity, calmness, fearlessness, positive-expectation, positive-fear, positive-hope, joy* (positive-emotion)
2. *negative-fear, sadness, general-dislike, ingratitude, shame, compassion, humility, despair, anxiety, daze* (negative-emotion)
3. *think, gravity, surprise, ambiguous -agitation, ambiguous-fear, pensiveness, ambiguous-expectation* (ambiguous-emotion)
4. *apathy, neutral-unconcern* (neutral-emotion)

This issue could be also faced by an ontology driven approach as in [14] and [18]. The related ongoing work is discussed in the last section.

The new set of emotions for *The Count of Montecristo* is:
`affection[1], anxiety[3], enthusiasm[1], general-dislike[1], joy[1], love[1], negative-fear[2]`

Then, a further reduction phase can be implemented associating the 32 emotions of the third level of WordNet-Affect to an extended set of Ekman emotions[11], formed by eight emotional categories *happiness*, *anger*, *disgust*, *fear*, *sadness*, *surprise*, *neutral*, *ambiguous*.
Taking in account these eight emotional categories, the set of emotions associated to the *The Count of Montecristo* is further reduced and includes:
`disgust[1], fear[5], happiness[4]`

5 Dataset

Dataset structure. The database provided by Zazie was constituted of 38374 records, each one representing the association of a tag in the MOOD set to a book by an user. Each record is represented by 7 fields (`user_id`, `book_isbn`, `book_title`, `book_pages`, `book_publisher`, `book_blurb`, `mood`).

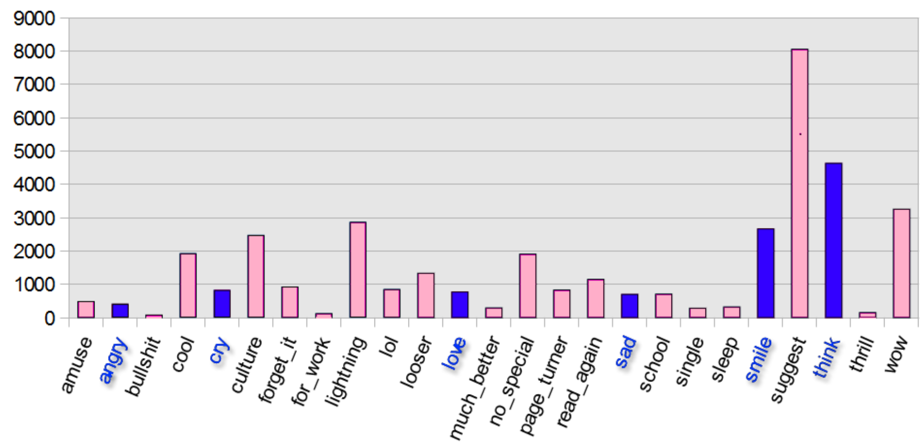

Fig. 2: Distribution of the records, with respect to the MOODs icons, in the Zazie database after the first filtering steps (19819 records)

Database filtering. The filtering phase has been implemented in five steps:

– Only the books that have received sufficient attention from the community have been selected. This criterion has been applied filtering all the books that received less than 5 tags i.e., only the books that appear in the database with at least 5 records have been kept. After this step the database contains 19819 records distributed as shown in Fig. 2.
– Records are grouped with respect to (`book_isbn`, `mood`) in order to compute how many users have chosen the *mood* tag for the book *book_isbn*. After this step the database contains 9644 records representing the books having the structure (`book_isbn`, `mood`, `#occ`).
– A filter based on the standard deviation values allows to select only the books associated with tags that are effectively representative. All the books having σ value lower than 1.5 were discarded.
– In order to avoid to associate a mood with too little occurrences, all those records having the tag frequency less than the arithmetic mean were discarded. After this phase the database contains 691 records.
– Only the records having MOOD value in the set M=\{*angry*, *cry*, *love*, *sad*, *smile*, *think*\} of the tags we are studying has been selected. After this step the database contains 236 records distributed as in Fig. 3

Although the initial database provided by Zazie developers was large, the dataset resulting by applying the techniques described in the previous sections is relatively small, containing 236 instances. Furthermore, the distribution among the classes is not uniform: there are two predominant classes (think with 95 instances and smile with 87 instances), and four minor classes with 54 instances. An ongoing work using a larger initial database and designing different filtering methods and rules is being implemented.

MOOD	#records	%records
angry	9	3,81%
sad	11	4,66%
love	14	5,93%
cry	20	8,47%
smile	87	36,86%
think	95	40,25%
	236	100,00%

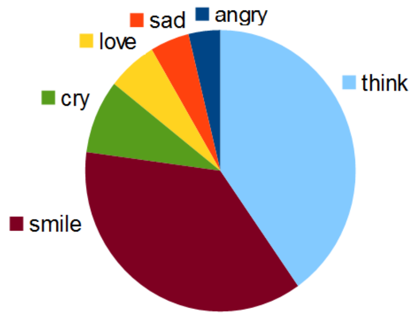

Fig. 3: Distribution of the records, with respect to the selected MOODs icons, in the Zazie database after all the filtering steps (236 records)

A series of preliminary tests was run in order to determine an appropriate value allowing to select a sufficient number of books where some tags were more popular than others, for example considering the arithmetic mean and the standard deviation of the MOODs frequency. It is important to note that this values depend on the specific dataset.

The following tables show an example of discarded and kept books; it is possible to note that, in table on the left, the frequencies for each tag are all similar, not allowing to determine a characteristic MOOD, while, in the table on the right, a book having a high standard deviation σ is shown. In this second case it is clear that the frequencies distribution is more heterogeneous and that the two most characteristic MOODs can be easily individuated.

ISBN	Mood	Freq	Mean	σ
978804548836	cool	2	1.75	0.50
978804548836	angry	2	1.75	0.50
978804548836	lightning	2	1.75	0.50
978804548836	lol	1	1.75	0.50

ISBN	Mood	Freq	Mean	σ
978806176556	cry	3	10.75	9.39
978806176556	angry	3	10.75	9.39
978806176556	culture	15	10.75	9.39
978806176556	think	22	10.75	9.39

Table 1: Database samples showing discarded (left) and kept (right) books

6 Experiments

Experiments were carried on in order to prove if an automated classification of book blurbs based on Zazie emotional tags is possible and can actually be used with a satisfactory accuracy. However other experiments and the design of other techniques to build a reliable dataset are ongoing works.

In this group of experiments the classes are identified by the selected MOODs $\{$*smile*, *love*, *sad*, *think*, *angry*, *cry*$\}$. Previous experiments presented in [23] considered the five most frequent MOODs i.e., *suggest*, *think*, *wow*, *smile* and *lightning*. The classification accuracy, showed in Table 2, was not satisfactory. A deeper analysis made clear that a motivation could lie in the meaning of the tags; in fact, tags as *wow*, *lightning* and *suggest* can not be related without ambiguity to an emotional term, in particular they can be used both for positive both for negative emotions, and so they do not seem to be suitable to be related to the emotional content of blurbs.

Among the information characterizing a book which is available in the Zazie database, the author and the emotions extracted by the blurb analysis have been used as the sample features. Different from [23], the publisher and the number of pages have been discarded, because they are associated to a particular edition of the book and do not characterize the book as its general literary work.

Therefore, in these experiments, each record in the dataset represents a book and is characterized by either 34 or 10 features:

- the author (nominal attribute)
- the emotions extracted from the blurb (numerical attributes) valued by their occurrences (32 or 8 depending on which strategy for emotion reduction is applied)
- the MOOD tag (nominal attribute) representing the class attribute

Note that further informations on the books are not used because our aim is to prove that an automated classification on Zazie is possible, with an acceptable accuracy, using only the information that is actually available in Zazie itself. An automated classification that uses other information and features, even if important, was out of our goal.

The experiments had been carried on by means of the software *Weka* [24], that supplies for the implementation of many machine learning algorithms and several measures for the model evaluation.

The experimentation has been realized through the *cross validation* technique with ten folds using, in particular, algorithms based on decision trees. Preliminary tests, also reported in [23], were run also using bayesian classifiers. Besides the good results it has demonstrated, the decision tree approach was preferred in this group of experiments because it returns a model (i.e. the tree) that is more readable and analyzable.

Models have been evaluated by the *accuracy*, *recall* and *precision* measures, defined as follow:

- $Accuracy = TP/N$ where TP is the number of instances correctly classified and N is the total number of instances.
- $Recall = \frac{1}{NC}\sum_{i=1..NC} Recall_i$ where NC is the number of classes, $Recall_i = \frac{TP_i}{TP_i+FN_i}$ and TP_i and FN_i are respectively the instances correctly classified as members of class i and the instances wrongly classified as not belonging to the class i.
- $Precision = \frac{1}{NC}\sum_{i=1..NC} Precision_i$ where $Precision_i = \frac{TP_i}{TP_i+FP_i}$ and FP_i is the number of instances wrongly classified as members of class i.

The results for *accuracy*, *recall* and *precision* for the dataset described in Section 5 are presented in Table 3 where comparisons among the used algorithms are shown: best accuracy levels are obtained with *J48* [25] and *BFTree*[27] algorithms that have equivalent performances, while, also in contrast with the results contained in Table 2, *LADTree* [26] and *RandomForest* [28] did not perform as well as expected. The analysis of motivations of these differences is ongoing but preliminary results show that algorithms having an unpruned version perform better. Moreover it seems, from the results, that there is not a great difference between the emotional model with 32 emotions (Tab.3a) and the one with 8 emotions (Tab.3b). It can be noted that, with respect to the results presented in Table 2, a significant improvement was obtained reaching more than 70% of accuracy.

Algorithm	**Accuracy**	**Recall**	**Precision**
`J48 -U -M 2`	30.46%	0.305	0.302
`BayesNet default`	35.08%	0.351	0.332
`LADTree -B 10`	44.96%	0.450	0.421

Table 2: Previous results: classes are identified by the five most frequent MOODs

(a) Emotional model with 32 emotions derived from WordNet-Affect

Attributes	**Accuracy**	**Recall**	**Precision**
`J48 -U -M 3`	70.31%	0.694	0.703
`BFTree -U -M 3`	72.92%	0.715	0.729
`LADTree -B 20`	58.85%	0.589	0.556
`RandomForest -depth 5`	66.15%	0.639	0.661

(b) Emotional model with 8 emotions derived from extended Ekman model

Attributes	**Accuracy**	**Recall**	**Precision**
`J48 -U -M 3`	72.02%	0.714	0.720
`BFTree -U -M 3`	70.46%	0.685	0.683
`LADTree -B 20`	52.85%	0.509	0.528
`RandomForest -depth 5`	68.91%	0.696	0.689

Table 3: Classification results with respect to selected emotional MOODs

7 Conclusions and Future Work

The aim of this work was to study the feasibility of an automated classification of books in Zazie according to the emotional tags by means of the lexical analysis of book blurbs. A supervised learning approach was used and an experimentation was implemented.

Although experiments were preliminary they are very encouraging, especially considering that improvements are expected from the ongoing works on database filtering and emotion extraction.

The blurb is confirmed to be a good source of emotional information about a book and it actually can be analyzed with the aim of sentiment analysis and emotion recognition. To the best of our knowledge this is the first attempt to apply sentiment analysis to books classification.

Further developments are split into different directions.

On the one hand an improved dataset has to be built: now some classes are not sufficiently represented and most of the misclassification errors arises for this reason.

Also the set of features characterizing the samples has to be extended and the approach presented in [15] using WordNet hypernyms and significant word is under implementation; furthermore an ontology driven approach that uses the ArsEmotica ontology presented in [18] is under consideration, to perform a different emotion extraction phase.

On the other hand, from the perspective of Zazie, a user feedback process could be implemented in order to confirm or contradict the MOOD chosen by the classifier and an emotion-driven search engine could be developed in Zazie.

After finishing off this promising but still preliminar and explorative approach, a more general classifier of books, which can process other domains different from Zazie's network, can be also a future goal.

Furthermore, given appropriate tools capable to collect information from the results of queries on general search engines (e.g., Google, Bing, Yahoo Search) or specialised repositories for books (e.g., Google Books, Amazon), a further application could be directed to the use in the preprocessing phase of the web based proximity measures analysed in [30] and [31] which can return the similarity between two or more words and have been already applied to semantics-driven search engines.

Acknowledgements

We are grateful to Marco Ghezzi and Zazie developers, Joe and David, for the collaboration to this research.

References

1. Y. E. Kim et al.: Music Emotion Recognition: a State of the Art Review. ISMIR 2010 - 11th International Society for Music Information Retrieval Conference, (2010)
2. Yang, Y.-H. and Chen, H. H.: Machine recognition of music emotion: A review. ACM Transactions on Intelligent Systems and Technology, 3(3), 1–30, (May 2012)
3. Habu Music. http://www.habumusic.com
4. Stereomood. http://www.stereomood.com
5. Barthet M., et al.: Multidisciplinary perspectives on music emotion recognition: Implications for content and context-based models. Proc. CMMR, 492–507, (2012)
6. Russell, J.A.: A circumplex model of affect. J. of Pers. and Social Psy. 39(6), 1161–1178 (1980)
7. Thayer, J.F.: Multiple indicators of affective responses to music. Dissert. Abst. Int., 47(12), (1986)
8. Gilad Mishne: Experiments with Mood Classification in Blog Posts. Style2005, Stylistic Analysis of Text for Information Access, (2005)

9. C. Ovesdotter Alm, at al.: Emotions from Text: Machine Learning for Text-based Emotion Prediction. Proc. of HLT and EMNL Conferences, 579–586, (2005)
10. Zazie. http://www.zazie.it
11. Paul Ekman: Facial Expression and Emotion. American Psychologist, 48(4), 384–392, (1993)
12. C. Strapparava, R. Mihalcea: Learning to Identify Emotions in Text. SAC, (2008)
13. E. Cambria, M. Grassi, A. Hussain and C. Havasi: Sentic Computing for social media marketing. Multimedia Tools and Applications, 59, 557–577, (2012)
14. M. Baldoni, et al.: From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. Intelligenza Artificiale 6(1): 41-54 (2012)
15. K. Roberts, M. A. Roach, J. Johnson, J. Guthrie, and S. M. Harabagiu: Empatweet: Annotating and detecting emotions on Twitter. In Proceedings of the LREC12, Istanbul, Turkey, pp. 3806–3813, (2012)
16. H. Liu, H. Lieberman and Ted Selker: A model of textual affect sensing using real-world knowledge. IUI 2003: 125-132
17. J. Suttles and Nancy Ide: Distant Supervision for Emotion Classification with Discrete Binary Values. CICLing (2): 121-136. (2013)
18. F. Bertola and V. Patti: Emotional Responses to Artworks in Online Collections. In Proceedings of PATCH 2013: Personal Access to Cultural Heritage, UMAP Workshop, volume 997 of CEUR Workshop Proceedings, (2013)
19. E. Pianta, L. Bentivogli and C. Girardi: MultiWordNet: Developing an aligned multilingual database. Proceedings of the 1st International WordNet Conference, 293–302, (2002)
20. George A. Miller: WordNet: a lexical database for English. Commun. ACM, 38(11), 39–41, (1995)
21. C. Strapparava and A. Valitutti, WordNet-Affect: an affective extension of WordNet. In Proc. of 4th International Conference on Language Resources and Evaluation (LREC 2004), 1083 – 1086, 2004
22. E. Zanchetta and M. Baroni: Morph-it! A free corpus-based morphological resource for the Italian language. Corpus Linguistics 2005, 1(1), University of Birmingham, Birmingham, UK (2005)
23. Fabiana Zollo: Classificazione automatica di libri rispetto ai tag emozionali del social network Zazie. Master Thesis, Department of Mathematics and Computer Science, Università degli Studi di Perugia, Italy (2013)
24. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and Ian H. Witten: The WEKA Data Mining Software: An Update. SIGKDD Explorations, 11(1), (2009)
25. Ross Quinlan: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, (1993)
26. G. Holmes, B. Pfahringer, R. Kirkby, E. Frank, and M. Hall: Multiclass alternating decision trees. ECML, Springer, 161–172, (2001)
27. J. Friedman, T. Hastie, R. Tibshirani: Additive logistic regression : A statistical view of boosting. In Annals of statistics. 28(2), 337–407, (2000).
28. Leo Breiman: Random Forests. In Machine Learning. 45(1), 5–32, (2001).
29. M. Grassi, E. Cambria, A. Hussain, and F. Piazza. Sentic web: A new paradigm for managing social media affective information. Cognitive Computation 3(3), pp. 480-489 (2011)
30. V. Franzoni, A. Milani: PMING Distance: A Collaborative Semantic Proximity Measure. WI–IAT, vol. 2, 442–449, IEEE/WIC/ACM (2012).
31. C. H. C. Leung, Y. Li, A. Milani, V. Franzoni: Collective Evolutionary Concept Distance Based Query Expansion for Effective Web Document Retrieval. Computational Science and Its Applications, 657–672 , Springer, (2013)