Dynamics of Emotions and Relations in a Facebook Group of Patients with Hidradenitis Suppurativa

Gianfranco Lombardo, Alberto Ferrari, Paolo Fornacciari, Monica Mordonini, Laura Sani, and Michele Tomaiuolo(✉)

Dipartimento di Ingegneria e Architettura, Università di Parma, Parma, Italy
michele.tomaiuolo@unipr.it

Abstract. Hidradenitis suppurativa (HS) is an orphan, underdiagnosed and painful disease of the skin that has a considerable negative impact on quality of life and on emotional well-being. As reported by the italian HS patients' association (Inversa Onlus), this condition brings patients to develop an emotional closure with the consequence that they often don't talk about their condition with anybody. In this paper we discuss some results obtained by applying automatic emotion detection and social network analysis techniques on the Facebook group of the Inversa Onlus association. In particular, we analyze the patients' emotional states, as expressed by the post published from 2010 to 2016, and how these emotions are influenced by friendships in the group, during the years.

Keywords: Social network analysis · Emotion detection
Sentiment analysis · Hidradenitis suppurativa · Facebook

1 Introduction

Hidradenitis suppurativa (HS) is an orphan and underdiagnosed disease of the skin that affects 1%–4% of the general population. The main characteristics of HS are chronic, inflammatory, painful boils in the folds of the skin that have a considerable negative impact on quality of life and on emotional well-being. It is widely reported that the difficulties in obtaining a diagnosis together with the lack of proper therapies bring HS patients to develop an emotional closure, with the consequence that patients often don't talk about their condition with anybody, as reported by the italian HS patients'association (Inversa Onlus) [1]. In the last few years, with the diffusion of social media, a considerable number of patients started to share a considerable amount of data related to their feelings. In this paper we present the results obtained using Sentiment Analysis on Inversa Onlus's Facebook group for the automatic detection of the emotional state of the patients from 2010 to 2016. In particular, this work focuses on the analysis of how friendships in the group influence the emotions expressed by patients

B. Guidi et al. (Eds.): GOODTECHS 2017, LNICST 233, pp. 269–278, 2018.
https://doi.org/10.1007/978-3-319-76111-4_27

over the years. Results have been obtained using a manual annotated training set and a seven-output hierarchical classifier system based on Parrot's emotion categorization (joy, love, surprise, fear, anger, sadness, and objective content) [2].

2 Related Works

Nowadays, online social networks provide more and more medical data; patients often share personal clinical information within an online community, with the aim of receiving emotional support. In recent years, the interest in patients' opinions and feelings expressed in web communities has considerably increased. One of the biggest challenges is to get a clear understanding of patients' condition. In [3] authors identify the 15 largest Facebook groups focused on diabetes management, analyzing 690 comments from wall posts written by 480 unique users. The work aims at identifying, with traditional manual method of content analysis, the main topics of the discussions. Since this approach is not scalable when trying to analyze huge quantities of data, it is necessary to introduce automatic analysis, based on machine learning algorithms. In [4] authors analyze different sources of patients' information (social media, blog, patients networks) in order to detect poor quality healthcare using sentiment analysis and natural language processing. Sentiment analysis using machine learning algorithms represents an automatic way to analyze sentiments, emotions and opinions from written language [5–7] and it's becoming increasingly important in the context of social media and for business and social sectors [8]. Our approach is innovative in joining two types of analysis:

- emotion detection applied to patients' Facebook posts, based on a hierarchical classifier using Parrot's emotion categorization (joy, love, surprise, fear, anger, sadness, and objective content) [2];
- analysis of the Facebook friendship relations among the members using Social Network Analysis, a discipline that focuses on the structural and topological features of the network, that can be useful in order to identify models of participation, opinion leaders and behaviors adoption [9].

As for the tools used, hierarchical classification is widely applied to big data collections [10–12] and in the scientific literature it is often considered an effective approach for emotion detection [13,14].

3 Methodology

In this section we present all the tools and algorithms we used in order to analyze and assign an emotion to the posts of each patient. According to Parrott's socio-psychological model, all human feelings could be partitioned with a primary categorization composed of six emotions, three positive (joy, love, surprise) and three negative (fear, sadness, anger). In order to obtain a performing system for emotion detection we used stopwords and stemming algorithms to preprocess the data, the bag-of-words model to extract features from the sentences, the Information Gain algorithm in order to reduce the features, and the Naive Bayes Multinomial classifier to optimize some hyperparameters in order to achieve better accuracy [15]. In fact, the whole system is built after a selection of the best mechanisms and parameters, in accordance to previous experiences [13]. Each step is described below, and the execution flow is presented in Fig. 1.

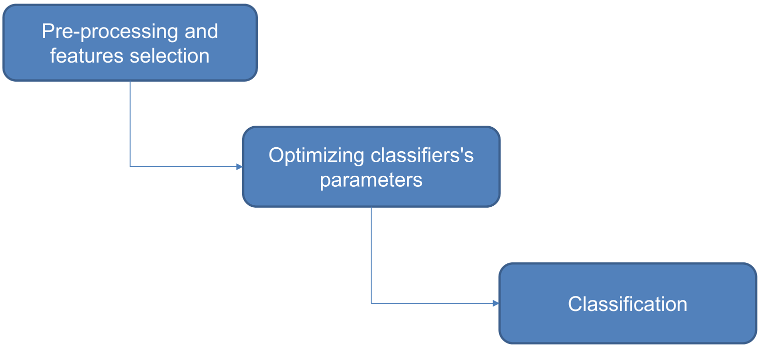

Fig. 1. Execution flow

3.1 Creating the Training Set

The first problem that we have considered has been the creation of a performing training set, to classify the patients' emotions. Due to the lack of useful datasets containing annotated posts of Italian patients, we have decided to create our own training set based on supervised learning and manual annotations. For creating the training set, we have considered the 10% of all the available posts, published in the Facebook group in the past seven years (Table 1).

Table 1. Number of posts in the training set for each emotion.

Sentiment	Number of posts in the training set
Objective	85
Love	70
Joy	80
Surprise	30
Fear	75
Sadness	85
Anger	65

3.2 Pre-processing and Features Selection

All posts have been pre-processed, in order to preserve only the elements with an emotional meaning. For this reason, we have used different automatic filters to remove Italian stopwords and punctuation, encode special characters, correct spelling mistakes, substitute contractions with their textual extension, and substitute smiles and emoticons with appropriate words. At the end of this process, sentences have been also filtered with a Stemming algorithm, in order to reduce inflected words to their word stem.

The last step has been the generation of features based on bag-of-words model using the StringToWordVector algorithm, turning each string into a set of attributes representing word occurrences. The TF-IDF function has been used to evaluate the relevance of each word in each sentence. Finally, these attributes have been filtered using the Information Gain algorithm, in order to extract the most meaningful features.

3.3 Classification

In this subsection we present the approach that we have used to build the resulting classifier. With the aim of building a hierarchical classifier, the manually annotated training set has been used to create four different training sets, in which each sentence is labeled according to the task of each classifier (Fig. 2):

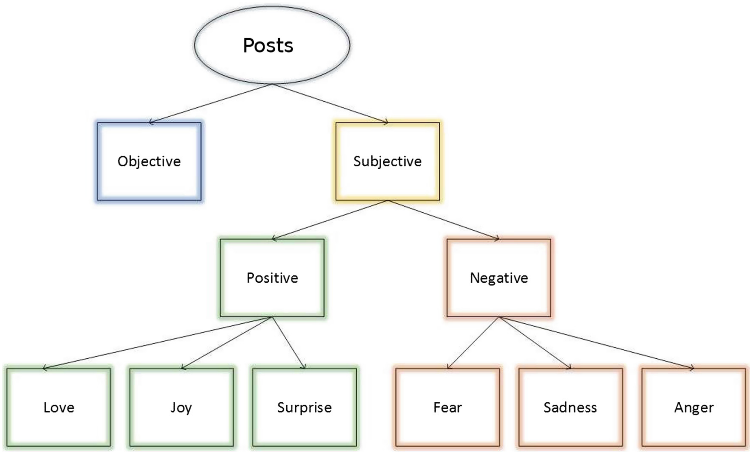

Fig. 2. Hierarchical classifier used

– **Objectivity/subjectivity classifier:** The training set used is simply the pre-processed training set in which every post associated with an emotion has been labelled as subjective while the remaining others, for example information requests or communications from the patient's association, have been labelled as objective.

– **Polarity classifier:** The training set used is the pre processed training set in which all of the posts previous labelled as subjective are divided into positive and negative.
– **Positive classifier:** The training set used is composed of the previous posts labelled as positive, divided into love, joy and surprise.
– **Negative classifier:** The training set used is composed of the previous posts labelled as negative, divided into anger, fear and sadness.

3.4 Optimizing Classifiers's Parameters

For each classifier used to build the seven-output hierarchical classification system, a preliminary analysis was performed. In this phase, we have selected the Naive Bayes Multinomial algorithm for implementing each classifier, as it produces the best results. Moreover, we have optimized some parameters, considered relevant for the training phase. In particular, we have searched at the same time for (*i*) the optimal length of N-grams to be used as features, and (*ii*) the number of features to select through the Information Gain algorithm. We have used a grid search, which is simply an exhaustive search through a manually specified subset of the parameters hyperspace of a learning algorithm, in order to select a grid of configurations using cross-validation to estimate the quality of classifiers configured according to them. Results are shown in Table 2.

Table 2. Parameters optimization results

Classifier	N-Gram (max)	Features
Sub/Obj	3	100
Pos/Neg	2	250
Fear/Sadness/Anger	3	340
Love/Joy/Surprise	2	100

4 Results

After having classified the emotions in patients' posts, in a period of seven years, we have performed further analysis. In this section we present the results, which may provide valuable hints for better understanding the Hidradenitis Suppurativa disease and its impact on patients' lives.

4.1 Distribution of Emotions

In this subsection we present the results obtained by analyzing the distribution of emotions for each month of the analyzed year. Looking at the distributions for each year (Fig. 3), and in particular the average distribution (Fig. 4), it is possible to note that the month in which patients are most active in the group is usually September, and to a lesser but considerable extent, January and April-May. These results could be read in different ways, but these three time frames

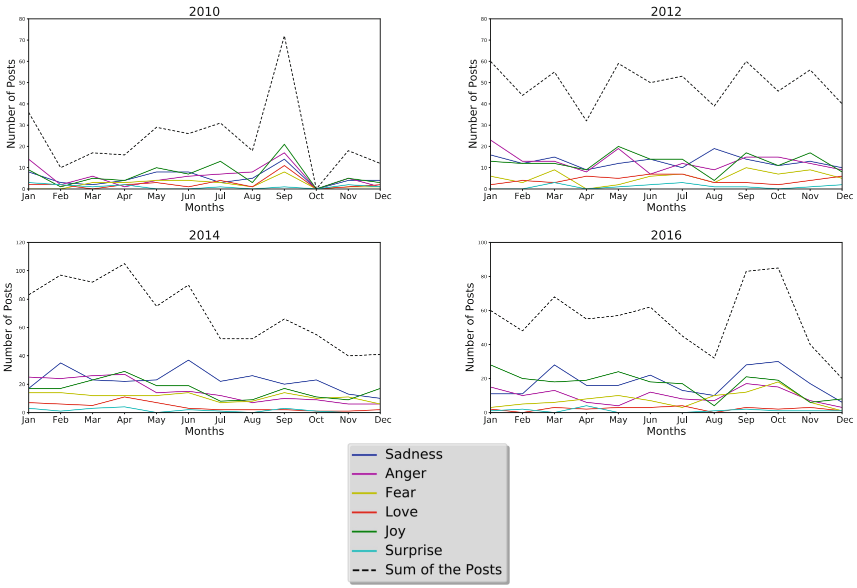

Fig. 3. Distribution of emotions in 2010, 2012, 2014, 2016.

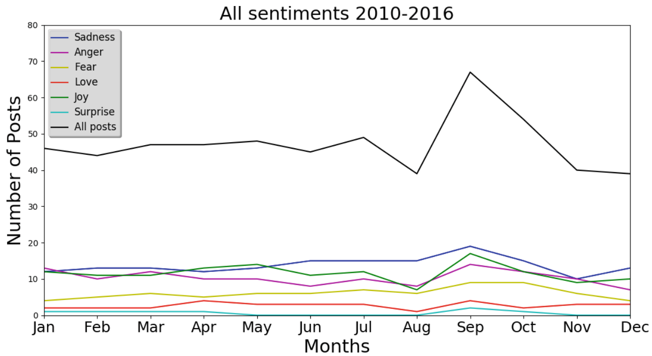

Fig. 4. Average distribution of emotions between 2010–2016.

have in common the fact that they follow holidays. In particular, the most active month (September) follows the traditionally long summer holidays. According to the negative impact of this disease [16], the prevailing emotion is sadness. Summing up the posts conveying negative emotions, they constitute the major component of the available posts. Despite that, the second prevailing emotion is joy, and in particular at the beginning of the group in 2010 it was the most present. This juxtaposition could be read in different ways, but it seems that in the group there are some influencers which express positive emotions in reaction to other negative posts. We will discuss this aspect later, analyzing the results concerning the social network analysis.

4.2 Social Network Analysis

After retrieving information about friendships on Facebook among the group members and classifying posts written by each patient in the group for each year, it is possible to analyze the established social network in the group and its evolution over the years (Fig. 5). In particular, at the beginning of the group activities, in 2010, there were few members, with many Facebook friendship relationships, expressing a variety of emotions. During the following years, the number of members in the group has constantly increased. At the same time, the differences about the emotions expressed by isolated members and connected members have grown.

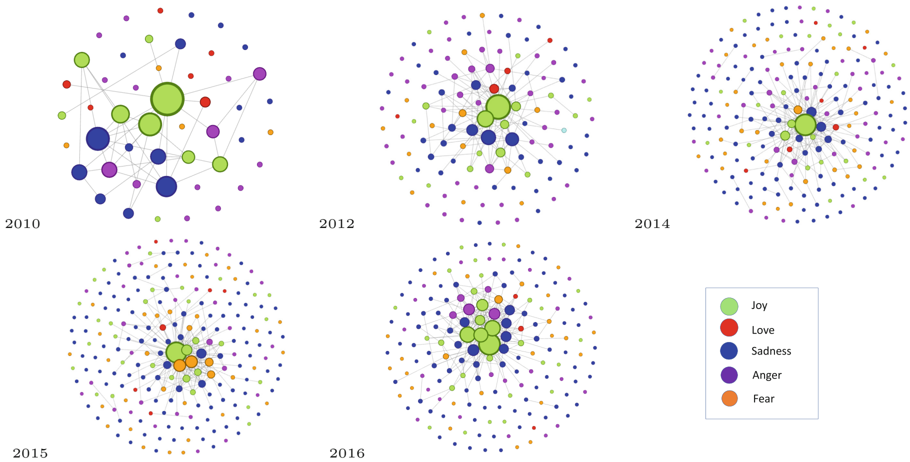

Fig. 5. Social network in 2010, 2012, 2014, 2015, 2016.

4.3 Friendship Relations and Expressed Emotions

The first result presented in this section is obtained analyzing how the emotions expressed by patients are correlated with their own degree in the social network (i.e., the number of friends in the group). Members are classified in four categories, according to the number of friendship relations in the group:

- **Zero Relationships**
- **Weak:** from 1 to 5 Relationships
- **Moderate:** from 6 to 19 Relationships
- **Strong:** 20 or more relationships.

The results obtained analyzing posts and social network in 2016 and in the period between 2010 and 2016 are presented in Fig. 6. It is worth noting that, for the nodes with higher degree, the predominant emotion is joy, while all the negative emotions decrease. As a possible interpretation, this result may represent an evidence of the positive influence of the group. Figure 6, for example, shows that all the patients that have developed many relationships with other peers in the group express positive emotions.

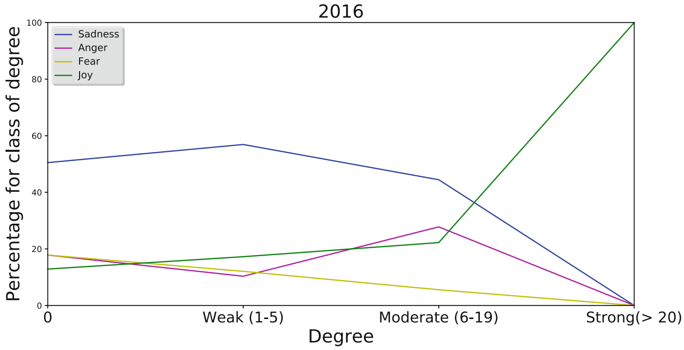

Fig. 6. Relationships between number of friendships and feelings expressed only in 2016

4.4 Changes of Emotions During the Years

Looking at the network between 2014 and 2016, it is clear that the major part of the active members don't have relationships with each others and they express mostly negative emotions. Despite that, analyzing in an anonymous way each patient, we have discovered that, in particular in the period from 2014 to 2016, members that used to express fear or anger emotions, but that were in connection with others, started to develop and express joy emotions. This result has been further analyzed, by calculating a matrix of transitions among emotions, considering the changes of emotions expressed by members with at least three relationships inside the group (Table 3). Observing the matrix, it is possible to note that transitions from anger or fear emotions to joy are more frequent than the transitions from sadness to joy. This result is important to understand that angry and worried members are more easily influenced by the rest of the social network. Instead, sadness appears to be the most static emotion. This essentially means that members who express sadness continue to express sadness, regardless of their own friendships in the group. At the same time, positive emotions as a whole are the most static class of emotions. In fact, transitions from positive to positive emotions are more frequent (53%) than transitions from positive to negative emotions (47%).

According to this multi-faceted analysis, it is thus possible to observe a significantly positive evolution of the group. However, while these analytical results were quite hoped by the group representatives, they were not obvious before this work. Thus, they also highlight the importance of the research and development of this kind of software tools. In future, it would be important to verify if this kind of social and psychological dynamics is common to similar online communities. In fact, to compare more data and possibly generalize our understanding of the phenomenon, we plan to continue this line of research and perform similar analysis in other online groups of patients, of different diseases.

Table 3. Matrix of transitions among emotions

	Joy	Love	Anger	Fear	Sadness
Joy	48%	11%	15%	4%	22%
Love	29%	0%	29%	13%	29%
Anger	29%	7%	14%	14%	36%
Fear	25%	17%	17%	8%	33%
Sadness	18%	3%	12%	6%	61%

5 Conclusion

The advent of social networks has led to new ways of expression for a support group composed of people who share the same disease or trouble. To date, it is still an open challenge of technology to provide tools suitable for a deepened analysis of these online communities and their influence over their participants. In this paper we have reported the results of an investigation of the dynamics of emotions and relations in a Facebook group of patients with hidradenitis suppurativa, who have been studied from 2010 to 2016.

We have detected the emotions of the members' posts by applying a hierarchical classification system, and then we have calculated the emotion of each member of the group and correlated it with his/her degree of relations in the group, that is the number of his/her friendship relations with other members of the community. This type of analysis has been performed for each year of the considered period.

The results have shown some interesting correlations among the decrease of negative emotions and the increase of the degree of relations among members. More generally, as the time progresses, it is possible to observe a positive dynamics of the members' emotions.

These results, though hoped, were not so obvious to the group representatives. In our opinion, such results highlight the importance of the research and development of this kind of software tools. In future researches, we plan to perform a similar analysis in other online groups of patients, of different diseases. This way, we will be able to compare the different results and verify if this kind of social and psychological dynamics is common to such online communities.

References

1. Bettoli, V., Pasquinucci, S., Caracciolo, S., Piccolo, D., Cazzaniga, S., Fantini, F., Binello, L., Pintori, G., Naldi, L.: The hidradenitis suppurativa patient journey in Italy: current status, unmet needs and opportunities. J. Eur. Acad. Dermatol. Venereol. **30**, 1965–1970 (2016)

2. Parrott, W.G.: Emotions In Social Psychology: Essential Readings. Psychology Press, Hove (2001)
3. Greene, J.A., Choudhry, N.K., Kilabuk, E., Shrank, W.H.: Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J. Gen. Intern. Med. **26**, 287–292 (2011)
4. Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., Donaldson, L.: Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual. Saf. **22**, 251–255 (2013)
5. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. **5**, 1–167 (2012)
6. Mohammad, S.M.: Sentiment analysis: detecting valence, emotions, and other affectual states from text. In: Emotion Measurement, pp. 201–238 (2015)
7. Kao, E.C.C., Liu, C.C., Yang, T.H., Hsieh, C.T., Soo, V.W.: Towards text-based emotion detection a survey and possible improvements. In: International Conference on Information Management and Engineering, 2009. ICIME 2009, pp. 70–74. IEEE (2009)
8. Franchi, E., Poggi, A., Tomaiuolo, M.: Social media for online collaboration in firms and organizations. Int. J. Inf. Syst. Model. Des. (IJISMD) **7**, 18–31 (2016)
9. Angiani, G., Fornacciari, P., Iotti, E., Mordonini, M., Tomaiuolo, M.: Models of participation in social networks. In: Social Media Performance Evaluation and Success Measurements, p. 196 (2016)
10. Dumais, S., Chen, H.: Hierarchical classification of web content. In: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 256–263. ACM (2000)
11. Silla Jr., C.N., Freitas, A.A.: A survey of hierarchical classification across different application domains. Data Min. Knowl. Discov. **22**, 31–72 (2011)
12. Addis, A., Armano, G., Vargiu, E.: A progressive filtering approach to hierarchical text categorization. Commun. SIWN **5**, 28–32 (2008)
13. Angiani, G., Cagnoni, S., Chuzhikova, N., Fornacciari, P., Mordonini, M., Tomaiuolo, M.: Flat and hierarchical classifiers for detecting emotion in tweets. In: Adorni, G., Cagnoni, S., Gori, M., Maratea, M. (eds.) AI*IA 2016. LNCS (LNAI), vol. 10037, pp. 51–64. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49130-1_5
14. Ghazi, D., Inkpen, D., Szpakowicz, S.: Hierarchical approach to emotion recognition and classification in texts. In: Farzindar, A., Kešelj, V. (eds.) AI 2010. LNCS (LNAI), vol. 6085, pp. 40–50. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13059-5_7
15. Chen, J., Huang, H., Tian, S., Qu, Y.: Feature selection for text classification with naïve bayes. Expert Syst. Appl. **36**, 5432–5435 (2009)
16. Onderdijk, A., Van der Zee, H., Esmann, S., Lophaven, S., Dufour, D., Jemec, G., Boer, J.: Depression in patients with hidradenitis suppurativa. J. Eur. Acad. Dermatol. Venereol. **27**, 473–478 (2013)