Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference

"Hang in There": Lexical and Visual Analysis to Identify Posts Warranting Empathetic Responses

Mimansa Jaiswal,[1] Sairam Tabibu,[2] Erik Cambria[3]
[1] Institute of Engineering and Technology [2] Indian Institute of Technology, BHU [3] Nanyang Technological University
mimansa.jaiswal@gmail.com, t.sairam.ece13@itbhu.ac.in, cambria@ntu.edu.sg

Abstract

In the past few years, social media has risen as a platform where people express and share personal incidences about abuse, violence and mental health issues. There is a need to pinpoint such posts and learn the kind of response expected. For this purpose, we understand the sentiment that a personal story elicits on different posts present on different social media sites, on the topics of abuse or mental health. In this paper, we propose a method supported by hand-crafted features to judge if the post requires an empathetic response. The model is trained upon posts from various web-pages and corresponding comments, on both the captions and the images. We were able to obtain 80% accuracy in tagging posts requiring empathetic responses.

Introduction

Empathy is often defined as the verbal or non-verbal gestures that evoke a sense of understanding of others' state of mind in a particular situation. Empathy encompasses several human interaction abilities, especially those that require the competence to reconstruct other person's words or actions and their perceived consequences.

Previous research has widely shown that agents without empathy are less preferred as compared to those who are empathetic, the latter being considered caring and likeable (Pestian et al. 2012). Empathy involves perspective taking, developing sensitivity to the other's affective state and communication of a feeling of care.

General health disclosures can be divided primarily into four categories:

(a) **Mental Health:** Issues related to stress, depression, feeling low, restless

"I am feeling low. I want to commit suicide"

(b) **Violence:** General acts of abusive behavior such as domestic violence, rape

"Today, I was raped"

(c) **Needing support:** Posts about losing a family member, tragedies which are temporal and not clinical

"I lost Pluto today. He was the sweetest dog I had ever known."

(d) **Physical Health:** Cases of physical discomfort such as sweating, pain, heart attack

"Please help me. I think I am having a heart attack."

Such cases of disclosures have been mapped and tracked through psychology. We refer to disclosures in the category 1, 2 and 3 as "*empathy-seekers*" thereafter.

The main contributions of the paper are as follows: (1) We propose a novel way to approach binary classification of empathy seekers. (2) We propose a generalized list of features that could work on different categories of posts. (3) We develop a standard corpus for empathy seekers (using search queries such as *soul-stirring, depression* from web-pages explained in the later categories)

The rest of the paper is structured as follows: Section II lists related works in empathy and affect; Section III presents the dataset development technique; Section IV provides the proposed method; Section V elucidates the experiments and results and finally, Section VI concludes the paper and offers pointers for future work.

Related Work

Prior research in psychology has examined the role of support from peers and society in combating mental health issues such as depression (George et al. 1989). Because our work forms a part of psycholinguistics, it has been demonstrated that the use of linguistic patterns can reveal important social and psychological aspects of an individual (Choudhary and De 2014). Previous works have dealt with use similar psycholinguistic cues to measure a specific area of health issue such as depression, suicide (Pestian et al. 2012) and bullying as well.

Dataset Development

The dataset was developed by sourcing images and captions with their respective response from various social media websites, namely tumblr, Facebook, Instagram and Buzzfeed. We skip twitter, which has been the natural choice for all social media analysis research, for its tendency to incline towards textual posts more than a combination of visual and textual, both. Also, on twitter, the distribution among retweets with response lies at a dismal 20%.

The main of dataset can be described as follows:

1) Develop a dataset which can be used to mark context that warrant an empathetic response.
2) Develop a dataset which can be used to identify empathetic and non-empathetic responses.

We utilize the Facebook API for storing posts and com ents from "Humans of New York" social page. The story is usually accompanied by a picture which usually relates to the caption. The second set of data is collected from image sharing websites such as Instagram and tumblr and the third one from Buzzfeed listicles. Because the listicles are usually a group of images, with common comments, we presently copy these comments, corresponding to all these images for response testing. Our final database comprised of 1000 anonymized context-response pairs of positive examples that we use for this study.

We add negative examples to our dataset by sourcing images related to happy events such as festivals, and by using search queries such as food, education and technology. To add non-empathetic responses, we ask ten people to reply to the post, as if they were trying to belittle the author or to hold them as a suspect. For the purpose of keeping our dataset unbiased, we thereby use inter-annotator agreement between four annotators to decide, if the post was marked correctly, i.e., ES/NES (Empathy-seeker or Non-Empathy seeker) and ER/NER (Empathetic response or Non-empathetic response).

The final dataset has the following distribution of context-response pairs: 330 of mental health issues, 283 of violence related issues and the remaining belonged to those requiring support.

Proposed Method

We model the task of empathy-seeker detection as a supervised classification problem in which each post is either classified as empathy-seeking or non-empathy seeking. We use six sets of lexical features and three sets of visual features to build our model. In the following subsections, we detail the features used and the classifiers that have been tried and compared.

Verbal Features

The verbal/textual features are used for two purposes: to classify the post as emotion seeker and to judge whether a response is empathetic or not (Cambria 2016).

Baseline Features

n-grams have been known as the best task-independent features for textual classification (Furnkranz 1998). Therefore, we choose n-grams as the baseline feature. We retrieved word n-grams, usually called as bag of words as bigrams and tri-grams and skip-grams (bi-grams) which after tf-idf transformation form our corpus. We filtered all the n-grams whose frequency was less than five, this set of feature would henceforth be called as baseline. These n-gram features are also used to identify temporal features such as *today, weeks etc* which are then used to identify the posts falling into the temporal issue category from the three categories we mentioned above. These temporal features have been known to be good linguistic attribute in identifying self-disclosure posts (Gibbs and Colston 2007).

Lexical Features

To model sentiment, we used emotional information from SenticNet (Cambria et al. 2016), a concept-level knowledge base for sentiment analysis that provides both semantic and affective information associated to words and multiword expressions by means of commonsense computing (Poria et al. 2013) and sentic computing (Cambria and Hussain 2015). SenticNet has been shown to model emotions such as satire (Poria et al. 2016), deception (Jaiswal, Tabibu and Bajpai 2016) and mood (Alam et al 2016) appropriately in previous research tasks, and hence we believe that it would present an appropriate representation.

Sentiment Amplification

As a general trend, it can be observed that almost all empathetic responses on social media make use of smileys, or specific punctuations. The use of quotation (“ ”)(Sander 1988) has been mentioned as an indication of inverse sentiment. Sentiment amplifiers have been used successfully to model ironical texts (Gibbs and Colston 2007), which is another form of emotion expression.

Speech act Features

A speech act has a performative function in the context of language and communication, i.e., it performs the function of apology, appreciation, gratitude etc. (Sander 1988). In our study, we use 7 kind of speech act features, as stated: apology, appreciation, response acknowledgment, opinioned response, non-opinioned response, gratitude, other.

We build a speech oriented classifier from SPAAC (Leech and Weisser 2003) using the above-mentioned features as to find the speech act distribution over our corpus.

Literary Device features

Hyperbole: Hyperbole is referred to as statements that tend to exaggerate the actual sentiment. This is usually mapped by occurrence of multiple positive or negative words consecutively (Sanders 1988).

Imagery: These are the words that create a visual understanding in mind of the reader. For example, "He took me to a close dark cabin", would be an example of imagery.

Psycholinguistic Features

To extract psycholinguistic features, we utilize the Linguistic Word Count (LIWC) (Pennebaker, Francis and Booth 2001), which is a knowledge based system that has been developed upon in the past decade. The utility of such features has been studied in various areas such as personality, age, deception, health. The types of LIWC features we use are:

General: word count, average words, word length
Psycholinguistic: affect, cognition
Personal concerns: work, achievement, home

Visual Features

Most of the websites allow addition of images along with posts. We use visual features to model personal images and know which of these warrant an empathetic response.

Facial presence

The first feature we use if based upon whether there is a presence of a face in the image or not. The model is run on both, the data from pages other than Humans of New York, and that combined. The feature vector models the presence of image, and if present, how many of them were there. We believe that self-focus extends to photographs too, while measuring isolation. We use an elementary face detection script based on an open source demonstration.

Gaze and facial sentiment

The second set of features took into account the gaze, if face was present, whether the participant was directly looking into the camera or away, and the facial angle from the vertical line. We also use OpenFace to measure Facial Action units and classify the sentiment projected by the face in the image, or average of the sentiments projected by the faces in the image. These three criteria were used to correlate introvert nature, social anxiety and isolation and hence form a part of our feature vector.

Hue and color

We take into account the image properties namely Hue, Saturation and Value. These three-color properties are commonly used in image analysis.

It has been observed that the happy individuals prefer vivid colors, while those feeling low or in need of support prefer darker colors (Carruthers, Morris, Tarrier and Whorwell 2010). We calculate pixel level averages to obtain HSV for our feature set, previously noted as satisfactory markers for mental health issues (Reece and Danforth 2016).

Features	LR	RF	LR+RF
Empathy Seekers Classification			
Verbal			
BF	65%	60%	70.2%
BF+LF+SA	66.23%	62.11%	73.03%
BF+LF+SA +LD	65.34%	63.72%	73.59%
BF+LF+SA+SF+LD+PF [a]	69.87%	65.93%	76.24%
Visual			
FP	58%	50.1%	70%
FP+ HSV	64%	61.2%	73%
FP+GFS	66%	61%	73.2%
FP+GFS+HSV [b]	68%	63.2%	74.33%
Verbal + Visual ([a]+[b])			
Mental Health (MH)	73.3%	69.40%	80%
Temporal Support (TS)	76.77%	69.78%	84%
Violence and Abuse (VA)	70.23%	65.18%	76.6%
MH + TS + VA	73.43%	68.12%	80.2%
Empathetic Response Classification (Only verbal features)			
BF	66.13%	63.2%	73%
BF+LF+SA	68.2%	64.55%	73.33%
BF+LF+SA +LD	69.87%	66.71%	75.6%
BF+LF+SA+SF+LD+PF [a]	72.16%	69.33%	78.9%

Table 1: Accuracy for combination of classifiers under various modalities

Experiments

We used three different classification methods to test the accuracy of our features, namely Logistic Regression (LR), Random forest (RF) and an ensemble of both of them. We perform an ensemble of LR and RF based on majority voting scheme. We use these two classifiers because they have the minimum relation amongst them, i.e., one models linear features while the other models non-linear ones.

We model ensemble classifier as follows:

*Ensemble_Classifier = w1*LR + w2*RF*
where w1+w2=1.0 and w1, w2 belong to {0.1, 0.9}.

We iterated over all possible combinations of w1 and w2 for the minimum cross entropy and settled upon 0.7 and 0.3 respectively. Individually, logistic regression produced the best result with an accuracy of over 76% with 99% confidence, while random classifiers averaged over 70% overall, probably due to overfitting on images. But a simple ensemble shoots the prediction accuracy of our model significantly, raising it up to 80.2% for overall classification and also 84% in some cases. Table 1 represents the f-scores of our model using different classification techniques with different feature sets on partitioned datasets.

Conclusion

We have proposed a method for identifying posts that require empathetic response. We have also tried to classify responses as empathetic or non-empathetic using a suite of classifiers.

Our model performs significantly well on classifying empathetic and non-empathetic responses, the f-scores averages to 79%, which, though cannot be compared to a benchmark due to lack of work in this area in the same context, beats the score of empathetic classification in call-center context of 70% (Alam, Danieli and Riccardi 2016). We believe this could be an important aspect in marking spam or hurtful responses or those that violate *Be Nice, Be Respectful* policy in social media forums.

We observe that ensemble classifiers perform the best and our use of gaze and HSV values in images combine with verbal features to give a superior performance. We believe that this performance would be enhanced if we took into account the photos that do not contain faces, but rather text or are stock images.

In future, we aim to deploy neural network for learning features other than our hand-crafted ones for they have been found to reduce the size of feature vector immensely (Poria et al. 2016).

References

Alam, F., Celli, F., Evgeny, A., Stepanov et al., 2016. The Social Mood of News: Self-reported Annotations to Design Automatic Mood Detection Systems. In: COLING, Osaka.

Alam, F., Danieli, M., and Riccardi, G., 2016. Can We Detect Speakers' Empathy? A Real-Life Case Study. In International Conference Cognitive InfoCommunications.

Cambria, E., 2016. Affective computing and sentiment analysis. IEEE Intelligent Systems 31(2), pp. 102-107.

Cambria, E. and Hussain, A., 2015. Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis. Cham, Switzerland: Springer, ISBN: 978-3-319-23654-4.

Cambria, E., Hussain, A., Havasi, C., and Eckl, C., 2009. Common sense computing: From the society of mind to digital intuition and beyond. In Biometric ID Management and Multimodal Communication, vol. 5707.

Cambria, E., Poria, S., Bajpai, R., and Schuller, B., 2016. SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In: COLING, 2666-2677.

Carruthers, HR, Morris, J., Tarrier, N., Whorwell, PJ, 2010. The Manchester Color Wheel: development of a novel way of identifying color choice and its validation in healthy, anxious and depressed individuals. BMC Med Res Methodol 10: 12.

Choudhary, M.D., De, S., 2014 .Mental Health Discourse on reddit: Self-disclosure, Social Support, and Anonymity. In proceedings of the Eight International AAAI Conference on Weblogs and Social Media.

Furnkranz, J., 1998. A study using n-gram features for text categorization. Austrian Research Institute for Artifical Intelligence, vol. 3, no. 1998, pp. 1–10.

George, LK, Blazer, DG, Hughes, DC, et al., 1989. Social support and the outcome of major depression. The British J. of Psychiatry 154(4), 478-485.

Gibbs, R., and Colston, H., 2007. Irony in Language and Thought: A Cognitive Science Reader, Lawrence Erlbaum Associates.

Jaiswal, M., Tabibu, S., Bajpai, R., 2016. The truth and nothing but the truth: Multimodal analysis for deception detection. In Proceedings, SENTIRE Workshop, ICDM.

Leech, G., and Weisser, M., 2003. Generic speech act annotation for taskoriented dialogues. In proceedings of the 2003 Corpus Linguistics Conference: Centre for Computer Corpus Research on Language Technical Papers.

Pennebaker, J.W., Francis, M.E., and Booth, R.J., 2001. Linguistic inquiry and word count: Liwc 2001. Mahway: Lawrence Erlbaum Associates, 71.

Pestian, J., Matykiewicz, P., Linn-Gust, M., et al., 2012. Sentiment analysis of suicide notes: A shared task. In proceedings of Biomedical informatics insights.

Poria, S., Cambria, E., Bajpai, R., and Hussain, A., 2017. A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion 37, pp. 98-125.

Poria, S., Cambria, E., and Gelbukh, A., 2016. Aspect extraction for opinion mining with a deep convolutional neural network, Knowledge-Based Systems, vol. 108, 42-49.

Poria, S., Cambria, E., Hazarika, D. and Vij, P., 2016. A deeper look into sarcastic tweets using deep convolutional neural networks. In: COLING, Osaka, 1601-1612.

Poria, S., Gelbukh, A., Agarwal, B., Cambria, E., Howard, N., 2013. Common sense knowledge based personality recognition from text. Advances in Soft Computing and Its Applications, 484-496.

Reece, A.G., Danforth, C., *Instagram photos reveal predictive markers of depression*, pre-print: arXiv:1608.03282v2.

Sanders, R. E., 1988. Dan sperber and deirdre wilson, relevance: Communication and cognition Language in Society, vol. 17, no. 04, pp. 604–609.