Twitter goes to the Doctor: Detecting Medical Tweets using Machine Learning and BERT

Kevin Roitero, Cristian Bozzato, Vincenzo Della Mea,
Stefano Mizzaro, and Giuseppe Serra
University of Udine, Italy.
{roitero.kevin},{bozzato.cristian}@spes.uniud.it
{dellamea.vincenzo},{mizzaro},{serra.giuseppe}@uniud.it

Abstract. We propose an effective model based on BERT to classify tweets as medical and non-medical. We experimentally validate the proposed model on more than 14k tweets, reaching accuracy levels of 0.93.

1 Introduction and Background

Twitter is a social media platform where million of users discuss and write on a daily bases about multiple topics. This large user base and the presence of available APIs makes Twitter a useful data repository for researchers. A research area that develops around Twitter consists in the categorisation of tweets, which allows to identify their topic [2, 8, 10, 5].

In this paper we propose to use machine learning models and in particular BERT [3] embeddings and MetaMap [1] to classify tweets as belonging to the medical or non-medical domain. We experimentally evaluate our approach on a dataset of more than 14k tweets, which we release to the research community.

2 Data

We collected the data used for the experiments as follows: we manually selected profiles of some sources (i.e., news websites, blogs, etc.) which publish articles that are categorised by the editors such that they include a "health / medical" category (or related ones). We considered the following sources of information: IFLScience, CNN, NBC News, PBS, USA Today, and BBC News (Science section). For such sources, we considered their official Twitter profile, and we considered only the tweets that included a full statement / article and an URL linking to the original domain; by exploring the categories on the original domain we where then able to discriminate between medical and non medical tweets. Let us make this process clear with an example. Let us suppose the IFLScience Twitter account publishes tweets with their respective URLs in the form iflscience.com/[topic]/[article_url]; IFLS uses as "health-and-medicine" as `topic` to identify medical related articles; thus, we consider such articles as belonging to the medical domain, and the others to do not. We adopt a similar approach for the other data sources.

By using such scraping strategy, we collected 14,582 tweets, 2095 labelled as being *medical* and the difference labelled as being *non-medical*. From such

data, we randomly extracted 500 medical and non-medical tweets as being our test set. The data used to conduct all the experiments can be downloaded at: https://github.com/KevinRoitero/twitterGoesToTheDoctor.

3 Methods

In this work we process the text of the tweets and we use it as a feature to predict the probability of a tweet as being part of the medical or non-medical domain. We consider the following machine learning models: *Logistic Regression* [9], which fits the data to a regression using a logistic function; *Random Forest* [4], an ensemble model based on decision trees classifiers; *Naive Bayes* [6], a probabilistic algorithm; and *Support Vector Machines* (SVM) [7], which places the target classes in a multidimensional space and separate them with an hyper-plane.

We feed such algorithms with two kind of features: starting from the text of the tweets, we extract *BERT* [3] embeddings, and *MetaMap* [1]terms. We consider both the cases of using BERT embeddings and MetaMap terms alone, or combining them together. BERT is an algorithm which computes the embeddings of a text considering the words in relation to all the other words in a context (e.g., a sentence). We use it to extract the embedding vectors of the text of the tweets in our dataset. Metamap is a well known tool for extraction of medical concepts from text. We use it to recognise, in Tweets, the presence of concepts belonging to one of the 127 MetaMap semantic types, hot-encoded. When we use the two techniques together we simply append the Metamap terms to the BERT embeddings.

BERT can be used both to extract embeddings or as a stand-alone classification algorithm; we consider both cases. When we use it as stand-alone classifier, we start from the pre-trained model released by Google, and we perform 2 epochs of fine-tuning on our training data.

4 Results and Conclusion

Table 1 shows the effectiveness scores of the considered algorithms; as in our setting we are more interested in correctly distinguish between medical and non medical tweets, we focus on the Accuracy and F1 measures. As we can see from the table, for both measures the BERT algorithm with no MetaMap terms achieves the highest scores; it is worth notice that also for the Precision and Recall metrics alone BERT is among the top performing algorithms, but the mos effective ones are Random Forest, Naive Bayes, and SVM. Furthermore, we see that while including BERT embeddings systematically provides an increase in effectiveness scores, MetaMap terms do not.

In conclusion, our contribution is twofold: we collect and release to the research community a set of tweets which can be used as a benchmark for tweet categorisation into medical and non medical, and we develop an effective classification algorithm based on machine learning and BERT. In future work we plan to extend such a technique to other domains.

Table 1: Effectiveness of the algorithms.

Base Model	Embeddings	Effectiveness Metrics			
		Accuracy	Precision	Recall	F1-Score
Logistic Regression	MetaMap	0.735	0.854	0.569	0.683
Logistic Regression	BERT	0.886	0.914	0.826	0.879
Logistic Regression	MetaMap + BERT	0.883	0.928	0.816	0.876
Random Forest	MetaMap	0.683	0.966	0.380	0.546
Random Forest	BERT	0.717	**0.982**	0.442	0.610
Random Forest	MetaMap + BERT	0.739	0.974	0.480	0.648
Naive Bayes	MetaMap	0.560	0.534	**0.955**	0.685
Naive Bayes	BERT	0.576	0.680	0.289	0.406
Naive Bayes	MetaMap + BERT	0.606	0.699	0.376	0.489
SVM	MetaMap	0.667	0.969	0.346	0.509
SVM	BERT	0.722	0.974	0.456	0.621
SVM	MetaMap + BERT	0.761	**0.982**	0.534	0.692
BERT	BERT	**0.929**	0.959	0.897	**0.927**
BERT	MetaMap + BERT	0.926	0.972	0.881	0.922

References

[1] Aronson, A.R., Lang, F.M.: An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 17(3), 229–236 (2010)

[2] Cotelo, J.M., Cruz, F.L., Enríquez, F., Troyano, J.: Tweet categorization by combining content and structural knowledge. Information Fusion 31, 54–64 (2016)

[3] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

[4] Liaw, A., Wiener, M., et al.: Classification and regression by randomforest. R news 2(3), 18–22 (2002)

[5] Quercia, D., Askham, H., Crowcroft, J.: Tweetlda: supervised topic classification and link prediction in twitter. In: Proceedings of the 4th Annual ACM Web Science Conference. pp. 247–250 (2012)

[6] Rish, I., et al.: An empirical study of the naive bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence. vol. 3, pp. 41–46 (2001)

[7] Suykens, J.A., Vandewalle, J.: Least squares support vector machine classifiers. Neural processing letters 9(3), 293–300 (1999)

[8] Tare, M., Gohokar, I., Sable, J., Paratwar, D., Wajgi, R.: Multi-class tweet categorization using map reduce paradigm. International Journal of Computer Trends and Technology (IJCTT) 9(2), 78–81 (2014)

[9] Wright, R.E.: Logistic regression. In: R.Yarnold, P., Grimm, L.G. (eds.) Reading and understanding multivariate statistics, p. 217–244. American Psychological Association (1995)

[10] Zhou, D., Chen, L., He, Y.: An unsupervised framework of exploring events on twitter: Filtering, extraction and categorization. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)