Polarity and Factuality Analysis of Medical Notes

Mohammad Amin Samadi
Electrical and Computer Engineering, Isfahan University of Technology
Isfahan, Iran

***Abstract*—**

Recent advancements in deep learning has led to major breakthrough in different areas within natural language processing (NLP) such as text classification, sentiment analysis and machine translation. In this work, the goal is to analyze messages on medical forums to answer certain questions regarding the sentiment underlying text and classify them into three classes with respect to polarity and factuality using machine learning and deep learning algorithms. Polarity consisted of three class: “positive”, “negative” and “neutral”; and factuality also predicts three classes: “opinion”, “experience” and “fact”. Two main approaches have been studied, machine learning using tf-idf feature extraction and deep learning models with word2vec input vectors. The latter proved to be more accurate when evaluated based on F1-measure, with a ten percent improvement.

Keywords*—*Deep Learning, Natural language processing, Sentiment Analysis

I. INTRODUCTION

Like any other machine learning task, natural language processing can benefit from unstructured data (e.g. in the form of text). Incorporation of unstructured data in machine learning models comes with its own challenges, such as requirement for more advanced data cleaning and preprocessing methods, as well as more powerful computational resources. However, considering the amount of valuable information that unstructured data holds, using them in action seems inevitable.

Medical domain of machine learning applications is no exception here as there have been more efforts recently to collect unstructured forms of data, e.g. medical reports, clinical text, making it also easier for doctors to record their thoughts on their judgement of patient conditions. In this work, we focus on using medical notes collected from a medical forum to facilitate medical decision-making. While clinical notes and medical reports hold information from the doctors’ or nurses’ point of view, forum-collected medical notes contain information from the patient’s perspective where patients try to describe how they are feeling or elaborating on their symptoms. Thus, sentiment analysis of such notes seems potentially very helpful.

People’s opinions are one of the most important sources of information. Information gathered from others’ experiences and knowledge can broaden vision and provide insight and perspective, and ultimately lead to better decision-making. Medical issues are no exception, as throughout history, people have been facing issues in dealing with medical problems and have struggled to overcome them fully. Although information obtained through people, especially in the medical field, are not always precise and accurate, it can help people gain a better sense of the condition they are facing and potentially guide them towards the right consultations or treatments.

This report is organized as follows. Section II provides information about the characteristics of the dataset used with Section III explains classification methods used. Experiment results and findings are presented in section IV. Sections V and VI wrap up this report with conclusions and potential future extensions.

II. DATASET

Data collected from medical forum Medhelp [12] via web scraping has been used where the training dataset [5] consists of 3600 sentences from forum replies labeled with respect to their polarity and factuality (see Fig 2). The distribution of labels across the two sentiment problems is available in Fig 1.

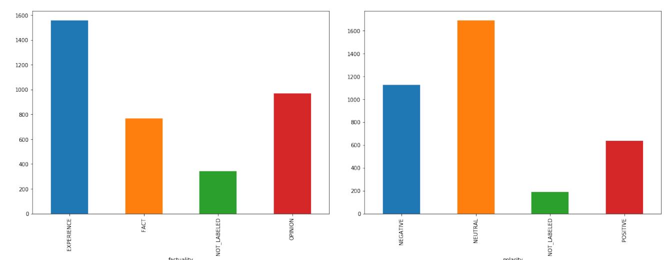

Fig. 1. Distribution of Data

link	title	label	text
https://medhelp.org/posts/Gastroenterology/Mor...	morning diarrhea occasional vomiting	Gastroenterology	morning diarrhea occasional vomiting anyone he...
https://medhelp.org/posts/Gastroenterology/HID...	hida	Gastroenterology	hida 17 year old daughter athlete upper abdomi...
https://medhelp.org/posts/Gastroenterology/aft...	poop wipe blood could	Gastroenterology	poop wipe blood could im 13 skinny wipe butt b...
https://medhelp.org/posts/Gastroenterology/Chr...	chronic bad taste	Gastroenterology	chronic bad taste long time complaining bad ta...
https://medhelp.org/posts/Gastroenterology/Bow...	bowel movement problems evacclean	Gastroenterology	bowel movement problems evacclean really need ...
https://medhelp.org/posts/Gastroenterology/dia...	diappearing gallstones	Gastroenterology	diappearing gallstones gallstones disappear co...
https://medhelp.org/posts/Gastroenterology/Bow...	bowel obstructions adhesions	Gastroenterology	bowel obstructions adhesions endured 3 bowel o...
https://medhelp.org/posts/Gastroenterology/emb...	embarrasing question	Gastroenterology	embarrasing question greetgings strange questi...
https://medhelp.org/posts/Gastroenterology/How...	lose weight ibs diarreah	Gastroenterology	lose weight ibs diarreah im 47 year old female...
https://medhelp.org/posts/Gastroenterology/utu...	utuuopuou	Gastroenterology	utuuopuou guhyiuojujtojkuuothis discussion rel...

text	polarity	factuality
I have been a coffee drinker on and off since ...	NEUTRAL	EXPERIENCE
I notice now that I am older, and perhaps I ha...	NEGATIVE	OPINION
When I drink it, if I drink about 2 cups a day...	NEGATIVE	EXPERIENCE
I always kind of said oh it's just allergies f...	NEGATIVE	OPINION
but in fact when I do not consume coffee, I am...	POSITIVE	EXPERIENCE

Fig. 2. Sample of dataset

III. CLASSIFICATION METHODS

A. Machine Learning

1) Tf-Idf Feature Extraction: In order to extract features from sentences, I have used the tf-idf [6] algorithm, which helps us find keywords in each class while ignoring the

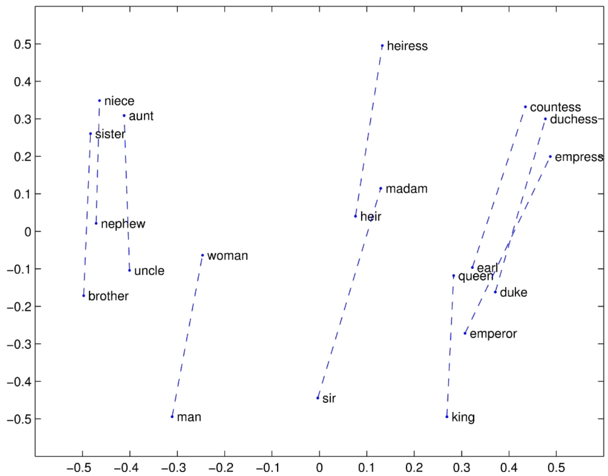

Fig. 3. Semantic relations in word embeddings

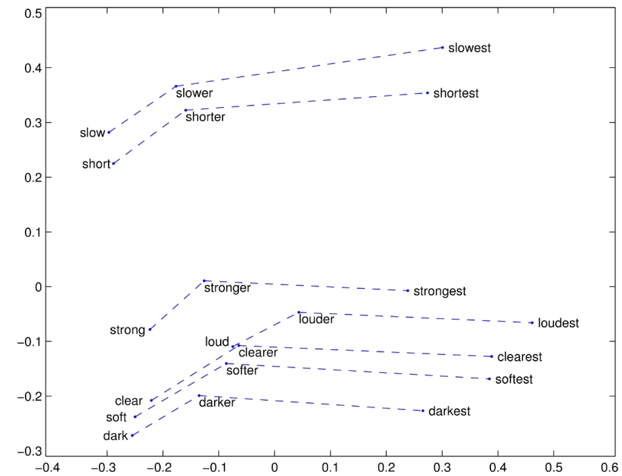

Fig. 4. Syntactic relations in word embeddings

frequent words that appear in most of the classes. By doing that, each sentence will be represented as a binary vector. Each element of that vector represents whether a specific word has occurred in that sentence or not. I have used these vectors as input to machine learning algorithms.

2) Machine Learning Algorithms: Machine learning algorithms usually tend to have faster training; however, fewer parameters may result in lower accuracy and F1 measure score for more complex data. I have chosen four algorithms that are believed to work better on text classification:

- Gaussian Naive Bayes
- Random Forest
- Online Passive Aggressive
- SVM

B. Deep Learning and Word2Vec

1) Word Embedding: Another approach to transform textual data into vectors is to represent each word with a vector. The most basic way to do so is to represent words as one-hot vectors. This technique has obvious disadvantages. Not only one-hot vectors do not convey any meaning relations, but also they are not memory efficient since size of the vectors is too big. A novel approach to represent words was introduced in [9] as the Word2vec algorithm. In word2vec, the main idea is that words which co-occur with each other more frequently should have vectors with more similarity than those that do not. By doing that, it is understood that vector representations can maintain valuable information such as semantic as syntactic relations between words. Some examples are available (see Fig 3 and Fig 4). To achieve this goal, a neural network with a hidden layer is trained with a fixed dimension, which is the dimension of word vector representation. The architecture of these algorithms is shown in Fig 5.

2) Deep Learning Models: In order to classify the sentences, I experimented with 3 deep learning architectures based on recurrent and convolutional neural networks. These three models are a deep bidirectional RNN model with LSTM cells in Fig 7, 1D CNN model inspired by [7] in Fig 6 and a layer of 1D added before a deep RNN for better feature extraction in Fig 8.

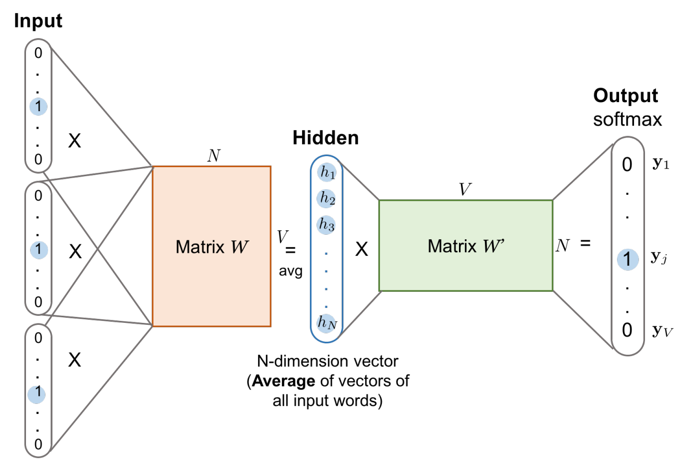

Fig. 5. Continous Bag-of-Words architecture

IV. EXPERIMENTS AND RESULTS

My goals in the experiments were to evaluate proposed models, and also to examine the influence of tf-idf parameters on machine learning classification performance based on f1-measure. To achieve these goals, I've evaluate machine learning models with different variations of parameters. Passive aggressive and bidirectional LSTM had the best performance among classifiers. Results are in Fig 9.

V. CONCLUSION

In this work, we studied sentiment analysis of medical notes using advanced methods in NLP. LSTM networks acheieved the best performance among various techniques, achieving precision, recall and F-1 score of above 0.7. The sequential characteristics of conversational forum data make LSTM networks a suitable algorithm for dealing with the complexities that come with human language. Distinguishing medical content in terms of the sentiment associated with them is the first, yet crucial, step towards next potential steps

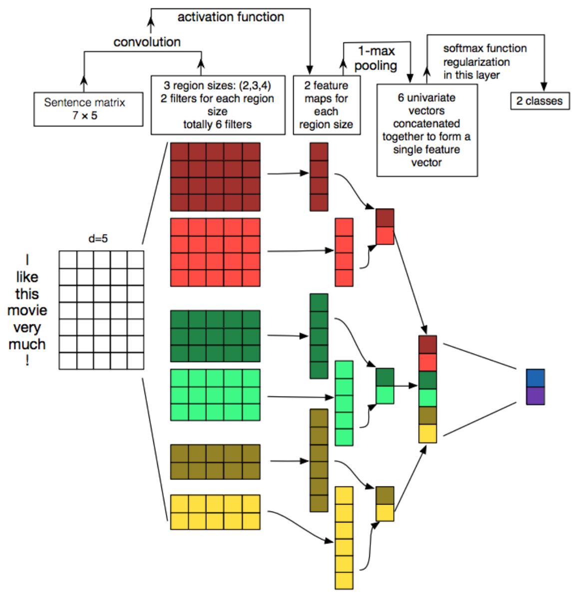

Fig. 6. Deep CNN architecture

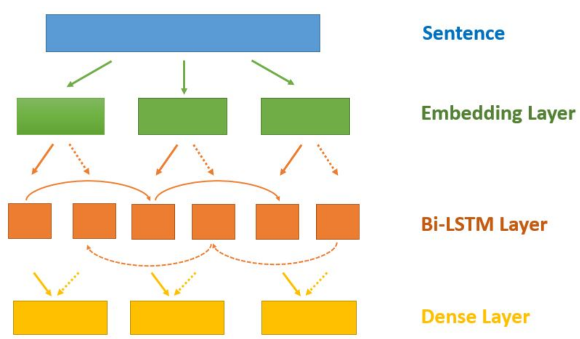

Fig. 7. Deep RNN architecture

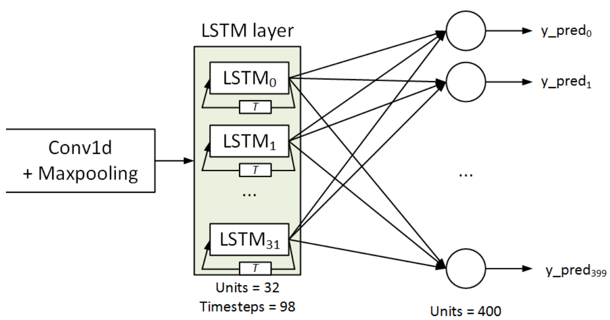

Fig. 8. 1-D CNN + Bi-LSTM architecture

in exploiting this data to perform preventive, diagnostic or prescriptive decision-making in medical domain, and LSTM networks seem promising in terms of this very first step.

VI. FUTURE WORK AND EXTENSION

Deep learning seems to be a powerful approach in sentiment analysis of medical data. However, such networks require word embeddings as their inputs. Although word2vec was shown to perform decently well in embeddings extraction, adding more context into these embeddings sounds like a natural next step in trying to push the performance further. Since this study is focused on a specific domain of data, i.e. medical, our prediction can benefit from incorporating domain-specific knowledge into our model. Thus, domain-specific word embeddings in the medical context is my next step as future work.

Model	Precision	Recall	F1-Score
Bidirectional LSTM	0.76	0.76	0.76
One Dimensional CNN + Bi-LSTM	0.74	0.75	0.74
One Dimensional CNN	0.74	0.75	0.74

		Gaussian NB			Passive Aggressive			Random Forest		
n-gram	features	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
1	5000	0.57	0.57	0.53	0.64	0.64	64	62	61	60
2		0.61	0.59	0.60	0.65	0.65	0.65	0.62	0.61	0.60
2	1000	0.52	0.44	0.44	0.61	0.60	0.59	0.58	0.58	0.58
	10000	0.61	0.59	0.59	0.66	0.66	0.66	0.63	0.63	0.62
	unlimited	0.61	0.59	0.60	0.67	0.65	0.63	0.60	0.59	0.58

Fig. 9. Performance comparison

REFERENCES

[1] Ali, T., Schramm, D., Sokolova, M., and Inkpen, D. Can i hear you? sentiment analysis on medical forums. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing* (2013), pp. 667–673.

[2] Baroni, M., Dinu, G., and Kruszewski, G. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (2014), vol. 1, pp. 238–247.

[3] Breiman, L. Random forests. *Machine learning 45*, 1 (2001), 5–32.

[4] Carrillo-de Albornoz, J., Rodríguez-Vidal, J., and Plaza, L. ediseases dataset, Nov. 2018.

[5] Carrillo-de Albornoz, J., Vidal, J. R., and Plaza, L. Feature engineering for sentiment analysis in e-health forums. *PloS one 13*, 11 (2018), e0207996.

[6] Joachims, T. A probabilistic analysis of the rocchio algorithm with tfidf for text categorization. Tech. rep., Carnegie-mellon univ pittsburgh pa dept of computer science, 1996.

[7] Kim, Y. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014).

[8] Koller, D., and Friedman, N. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[9] Mikolov, T., Chen, K., Corrado, G., and Dean, J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[10] Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., and Xu, B. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (2016), vol. 2, pp. 207–212.

[11] word2vec by tesnsorflow from https://www.tensorflow.org/tutorials/representation/word2vec

[12] www.medhelp.org