Medical Causation In Defining Emotions

Gulnara Bektemyssova and Aidos Sabdenov

May 20, 2021

Medical Causation In Defining Emotions

Bektemyssova Gulnara and Sandemov Aidos

[1] Candidate of technical sciences, associate professor of the department of Computer Engineering and Information Security, Processing homogeneous and heterogeneous data, International Information Technology University, st. Manasa 34/1, Almaty, Kazakhstan. g.bektemisova@iitu.edu.kz

[2] Phd, International Information Technology University, st. Manasa 34/1, Almaty, Kazakhstan.
24650@iitu.edu.kz

Abstract. Emotions are an essential component of human nature, which can describe a person's health and help determine this condition's causes. Proceeding from this, it becomes obvious that health plays a vital role in forming one of emotion condition, and in the reverse order, any emotion can describe the state of human health. This approach can provide medical personnel with important information about patients: emotions, state of health, and establishing cause and effect relationships. However, the creation of this model is hampered by the lack of large labeled datasets. Thus, the study's main goal is to create a dataset that would have information about the emotional state of a person and causal medical relationships that affect a person's emotional state. We conduct comprehensive data collection and analysis, using state-of-the-art models for assessing emotions, medical extractions of creatures, and determination of cause-and-effect relationships.

Keywords: eHealth · Emotion recognition · Named Entity Recognition · cause-effect

1 Introduction

With online communication progress, emotional information becomes significantly valuable not only for social research but also for medical analysis. Life-threatening, severe symptoms such as coughing, breathing difficulties, heart failure, and fatigue cause a compassionate person's state, leading to various feelings and emotions, such as surpriser to anger or fear to joy, and others. Given emotions, for example, help detect treatment effect and state condition of human.
There are several problems in the research of emotion cause extraction. The most notable is no data for emotion cause analysis. First studies defined as an emotion cause extraction problem described in Lee et al. (2010) [1]. Studying the experience of given research Gui et al. (2016) [2] improved and released a novel dataset that becomes a benchmark dataset for emotion cause extraction research. The emotion cause extraction task was also studied in novel researches where the problem was addressed as a clause-level binary classification problem

(Li et al., 2018; Xu et al., 2019; Yu et al., 2019) [3–5]. The next problem stands for the small size of the annotated corpus. Consequently, many deep learning models are not relevant for emotion cause extraction. The last problem in our research is defining the relationship between causes and health. Up-to-date improvement of medical text extraction researches (Habibi et al., 2017; Bhasuran and Natarajan, 2018; Giorgi and Bader, 2018; Wang et al., 2018; Lim and Kang, 2018; Yoon et al., 2019) [6–11] was made possible by applying machine learning techniques for medical named entity recognition (NER) and relation extraction (RE) applying modern models as Conditional Random Field Long Short-Term Memory. However, extracting medical text mining has limitations. To tackle this problem, recent study BioBERT (Lee et al., 2019) [12] outperform all previous work and become a state-of-the-art benchmark for NER and RE tasks, which is based on powerful model BERT (Devlin et al., 2019) [13].
The paper has the following organization. Section 2 stands for the motivation for our research. While section 3 discusses the novel corpus creation, including algorithmic and implementation terms. Section 4 reviews the results, and Section 5 concludes and discus future work.

2 MOTIVATION

Recent improvements in NLP, also known as the Spacy, BERT, and Spark NLP, give essential improvements in many state-of-the-art NLP benchmarks, such as sentiment analysis, question answering, and named entity recognition using deep learning. Deep learning opens an extensive range for research in any field and complex field as medicine. However, to work with deep learning, large data is required. Our work aims to build a model for analyzing human emotional behavior according to medical and other causes of this emotion. The main problem we tackled is the lack of quality data. For this purpose, we decide to create a corpus for our future research based on this topic.

3 Construction of Emotion Cause Corpus

In this section, we first describe the linguistic phenomenon in emotion expressions. It serves as the inspiration to develop the annotated dataset. We then introduce details of the annotation scheme, followed by the construction of the dataset. Today there is a lack of research and dataset for Emotion Cause, which makes this work relevant. To date, there are two studies by Xia Ding (2019) and Ghazi at al. (2015) [14, 15].

3.1 Emotion-Cause Pair Extraction Corpus ECPE corpus was constructed based on ECE corpus (Gui et al., 2016), where one utterance belong to one emotion and related to one causes. ECE corpus consist of Chinese news containing 20,000 articles. After removing irrelevant instances, there are 2,105 instances with cause relation. Emotion cause annotated as <cause >, and the emotion as <keywords >. Where, 97.2% of data has one emotion cause, other 2.8% respectively.

Example from data:
with cause:
<keywords >sadness <keywords >, <cause >sadness <cause >, because of sadness excessive
without cause:
<keywords >fear <keywords >, <cause><cause >, there are lingering palpitations, she still has lingering palpitations

3.2. Emotion-stimulus data The Emotion-stimulus data corpus [16] consist of 820 row of data including emotion tag and emotion cause. Data annotated in XML format: <cause>and <\cause>belongs to emotion cause. However, <emotion type>describes emotion. The given study was built with FrameNet tool (Fillmore et al.,2003) [17] into the frame of Ekman's six emotion classes (Ekman, 1992) [18] and finally annotated my human to verify them.

Example from data:
<fear >People are becoming more and more concerned , <cause >about the healthiness of their diet and way of life <\cause >. <\fear >

3.3. Custom Web Dataset Despite the fact that data were collected from available corpuses, this is still insufficient for extensive analysis. As a result, additional data was collected from "Psychiatric Treatment Adverse Reactions" (PsyTAR) Zolnoori (2019) dataset [20] and medical forums with total amount - 4000 of data. The difficulty lies in the fact that they are not annotated for causal relationships. This is the main task from which the following steps stand out. We split the emotion cause extraction task into two subtasks with the purpose to get a set of emotion clauses:

$$E = \{c_1^e, ..., c_n^e\} \tag{1}$$

and set of cause clauses for each document.

$$C = \{c_1^c, ..., c_n^c\} \tag{2}$$

For cause relevance, we decide to use a keyword matching pattern. According to Lee et al. (2010) there can be distinguished six linguistic groups of keywords that are essentially correlated with causes, as shown in Table 1. By given keywords, the manual corpus will be filtered.

For emotion relevance we collected data from two datasets: Twitter Emotion Corpus (TEC) [21] and CrowdFlower (CF) [**?**] with total amoubt: 61051 tweets and trained it on four models: Naive Bayes (NB) and Support Vector Machine (SVM), BERT, Multi-label BERT

Table 1. Key words for cause detection.

Group	Key Words
I:Prepositions	‘for’, ’as’
II:Conjunctions	‘because’, ‘so’, ‘but’, ’after’
IV:ReportedVerbs	‘to think about’, ‘to talk about’
V:EpistemicMarkers	‘to hear’, ‘to see’, ‘to know’, ‘to exist’
VI:Others	‘is’, say’, ‘at’, ‘can’

After prepossessing, the number of examples per emotion decreased significantly, due to significant noises in data. As a result, we manually picked from filtered data 800 examples for each emotion for training and 200 examples for emotions for the test set.
We use the F1 score for evaluation, which calculated:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{4}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{5}$$

After training on given models we got we that multi label BERT outperforms different model. For other models results described in Table 2 :

Table 2. Emotion classification results

Model	NB	SVM	BERT	BERT$_{ml}$
F1 -score	85%	87%	92%	96%

As a result, 96% of F1 - score was obtained, which is a best result for this task.

Example from data:
<happy >I feel much better after <cause >taking the headache medicine. <\cause ><\happy >

4 4. RESULTS

As a result of three corpora, we got a single dataset consisting of ECPE, ESD, and CWD, described in Table 3. But since the main goal is to identify the medical reason in a particular emotion, medical relevance will be applied to the

Table 3. Total number of sentence by emotion.

Emotion	ECPE	ESD	CWD	Sum
Anger	302	168	145	615
Disgust	225	40	106	371
Fear	379	129	104	612
Happiness	544	211	127	882
Sadness	567	98	113	778
Surprise	88	53	101	242
Total	2105	699	696	3500

assembled dataset.

For medical relevance we decide to use BioBERT, which significantly outper-forms previous state of the art researches in different types of medical text miningtasks, such as question answering (MRR by 12.24%), named entity recognition(F1 by 0.62%) and medical relation extraction (F1 by 2.80%) Medical clauses

$$M = \{c_1^m, ..., c_n^m\} \tag{6}$$

After applying medical relevance we got final results R as subtraction of given sets:

$$R = \{c_1^e, ..., c_n^e\} - \{c_1^c, ..., c_n^c\} - \{c_1^m, ..., c_n^m\} \tag{7}$$

The amount of data that is related to medicine decreased from 3500 to 986 data units, which is about 28% of all data. Final annotated data have XML format annotation. Where, <cause>and <\cause>belongs to the emotion cause. However, <mcause>and <\mcause>to the medical-emotion. For emotion <emotion type>tag was applied.

Example from data:
For medical cause:
<happy >I feel much better after <mcause >taking the headache medicine. <\mcause ><\happy >
For other cause:
<sad >I am really sad <cause >nobody wants to do it like I have done it for them. <\cause ><\sad >

5 CONCLUSION and DISCUSSION

In this paper, we present our work on medical causation in defining emotions. Lack of data for building and training a model was the driving force for creating corpus. We also describe the medical emotion cause extraction method to capture required data consisting of 3 main methods: emotion relevance, cause relevance, medical relevance. For emotion and medical relevance, state-of-the-art BERT models were used. However, cause relevance stands for the key word matching method, which needs improvement in future work. Given corpus helps us create the first model for analyzing and extracting emotional causes related to health and different events. We believe that the proposed work will help better investigate treatment effect and help understand human health's real state.

References

1. Chen,Y., Lee,S., Li.S, and Huang,C.: Emotion cause detection with linguistic constructions. In Proceedings of the 23rd International Conference on Computational Linguistics (COLING), pages 179—187 (2010)
2. Gui,L., Wu,D., Xu,R., Lu,Q., and Zhou,Y.: Event-driven emotion cause extraction with corpus construction. In Empirical Methods in Natural Language Processing (EMNLP), pages 1639—1649 (2016)
3. Li,x., Song,K., Feng,S., Wang,D., and Zhang,Y.: A co-attention neural network model for emotion cause analysis with emotional context awareness. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), 4752—4757 (2018)
4. Xu,B., Lin,H., Lin,Y., Diao,Y., Yang,L., and Xu,L.:Extracting emotion causes using learning to rank methods from an information retrieval perspective. IEEE Access. (2019)
5. Yu,X., Rong,W., Zhang,Z., Ouyang,Y, and Xiong,Z.:Multiple level hierarchical network-based clause selection for emotion cause extraction. IEEE Access, 9071—9079. (2019)
6. Habibi,M. et al.: Deep learning with word embeddings improves biomedical named entity recognition. Bioinformatics, 33, 37—48 (2017)
7. Bhasuran,B. and Natarajan,J.: Automatic extraction of gene-disease associations from literature using joint ensemble learning. PLoS One, 13, e0200699 (2018)
8. Giorgi,J.M. and Bader,G.D.: Transfer learning for biomedical named entity recognition with neural networks. Bioinformatics, 34, 4087 (2018)
9. Wang,X. et al.: Cross-type biomedical named entity recognition with deep multi-task learning. Bioinformatics, 35, 1745—1752 (2018)
10. Lim,S. and Kang,J. Chemical–gene relation extraction using recursive neural network. Database. (2018)
11. Yoon,W. et al.: Collabonet: collaboration of deep neural networks for biomedical named entity recognition. BMC Bioinformatics, 20, 249 (2019)
12. Lee, J., Yoon,W., Kim, S., Kim,D., Kim,S., So,C., Kang,J.: BioBERT: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics, Volume 36,1234–1240 (2020)
13. Devlin,J. et al.: Bert: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA. pp. 4171—4186 (2019)
14. Xia,R. and Ding,Z.:Emotion-cause pair extraction: A new task to emotion analysis in texts. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1003-–1012 (2019)

15. Diman Ghazi, Diana Inkpen, and Stan Szpakowicz.: Detecting emotion stimuli in emotion-bearing sentences. In International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), pages 152—165 (2015)
16. Ghazi,W., Inkpen,D. and Szpakowicz,S.:Computational Linguistics and Intelligent Text Processing Lecture Notes in Computer Science, Vol. 9042, 152–165 (2015)
17. Fillmore, C.J., Petruck, M.R., Ruppenhofer, J., Wright, A.: FrameNet in Action: The Case of Attaching. IJL 16(3), 297-–332 (2003)
18. Ekman, P.: An argument for basic emotions. Cognition Emotion 6(3), 169-–200 (1992)
19. Zolnoori, M. et al.: "The PsyTAR dataset: From patients generated narratives to a corpus of adverse drug events and effectiveness of psychiatric medications." Data in Brief 24. (2019)
20. Twitter Emotion Corpus 2012. http://saifmohammad.com/WebPages/SentimentEmotionLabeledData.html
21. CrowdFlower. 2016. https://www.figureeight.com/data/sentiment-analysis-emotion-text/.