Virus Causes Flu: Identifying Causality in the Biomedical Domain using an Ensemble Approach with Target-specific Semantic Embeddings

Raksha Sharma[1][0000−0003−2905−0194] and Girish Palshikar[2][0000−0003−3625−6705]

[1] Indian Institute of Technology, Roorkee, India
raksha.sharma@cs.iitr.ac.in
www.rakshasharma.com
[2] TCS Research, Tata Consultancy Services, Pune, India
gk.palshikar@tcs.com
www.tcs.com

Abstract. Identification of Cause-Effect (CE) relation is crucial for creating a scientific knowledge-base and facilitate question-answering in the biomedical domain. An example sentence having CE relation in the biomedical domain (precisely Leukemia) is: *viability of THP-1 cells was inhibited by COR*. Here, *COR* is the cause argument, *viability of THP-1 cells* is the effect argument and *inhibited* is the trigger word creating a causal scenario. Notably CE relation has a temporal order between *cause* and *effect* arguments. In this paper, we harness this property and hypothesize that the temporal order of CE relation can be captured well by the Long Short Term Memory (LSTM) network with independently obtained semantic embeddings of words trained on the targeted disease data. These focused semantic embeddings of words overcome the labeled data requirement of the LSTM network. We extensively validate our hypothesis using three types of word embeddings, *viz., GloVe, PubMed*, and *target-specific* where the target (focus) is Leukemia. We obtain a statistically significant improvement in the performance with LSTM using GloVe and target-specific embeddings over other baseline models. Furthermore, we show that an ensemble of LSTM models gives a significant improvement ($\sim 3\%$) over the individual models as per the t-test. Our CE relation classification system's results generate a knowledge-base of 277478 CE relation mentions using a rule-based approach.

Keywords: Cause-Effect Relation Extraction · Biomedical Domain · Deep Neural Network (Long Short Term Memory) · Semantic Embeddings.

1 Introduction

The MEDLINE database is growing at the rate of $500,000$ new citations each year. With such explosive growth, it is challenging to keep up to date with all of the discoveries and theories in biomedical research. Thus, there is a need to

provide automatic extraction of the user-oriented biomedical knowledge [1, 4]. Cause-Effect (CE) relation is one such type of user-oriented biomedical knowledge. The semantic connection between a causal argument and its effect is referred to as a CE relation. For example, *virus causes flu* has a CE relation, where *virus* is the cause argument, and *flu* is the effect argument, and *causes* is the trigger argument creating causal relation. Moldovan et al., [19] reported that causal questions are answered with a very low precision score of 3.1%. It is crucial to answering causal questions with high precision in the biomedical domain as it is related to human life. Identifying CE relation from the biomedical data can produce a scientific knowledge-base that can facilitate answering user queries in the biomedical domain [8]. The following example illustrates the purpose of the identification of CE relation in the biomedical domain.

- Input: "Tumor cell killing was achieved by concerted action of necrosis apoptosis induction."

- Proposed Output: CE relation found with the following CE mentions:

  Causal Cue: *achieved by*

  Cause: *Concerted action of necrosis apoptosis induction*

  Effect: *Tumor cell killing*

- QA System based on the proposed output:

  Question: What is the effect of concerted action of necrosis apoptosis induction on tumor cells of Leukemia?

  Answer: Tumor cell killing

The correct answer to the question could help understanding the disease to the patient or diagnosing a terminal illness such as *Leukemia* to the doctors/patients. Utilizing cause-effect relations in the development of a question answering system leads to improved performance [8].[3] Another direct application domain is a scientific database dedicated to a disease. Record of arguments of CE relations for a disease *viz., cause, effect, the cue for causality* can form a scientific knowledge-base dedicated to the disease [23]. Such knowledge-bases can help scientists, doctors, and other users perform tasks such as diagnosis, exploring and validating hypotheses, understanding the state-of-the-art, and identifying opportunities for new research.

Various complex constructs are used to express causality in text. The simplest way of expressing CE relations in the text is by using generic *causative* verbs,

[3] Causal questions are frequently used in general on Web. Naver Knowledge iN, http://kin.naver.com reported 130,000 causal questions from 950,000 sentence-sized database [19].

such as *cause, lead, result*. Apart from this, different domains have their causative verbs, which are either new verbs specific to that domain (*e.g.*, *over-express, up-regulate* in the biomedical domain) or generic verbs that have a special causative sense specific to that domain (*e.g.*, *inhibit, express* in the biomedical domain). There are other complexities with the linguistic expression of CE relations in text. One is the negation of the apparent CE relation mention, *e.g.*, *However, the precise mechanisms by which BCR stimulation leads to accumulation of malignant cells remain incompletely understood.* Next is the use of discourse connectives like *after, while etc.,* to express causal linking between two arguments, *e.g.*, $\{$*Cleaved caspase-3 was increased*$\}_{Effect}$ *after* $\{$*treatment of COR*$\}_{Cause}$. The presence of linguistically complex constructs in the biomedical domain makes extraction of CE relation a more challenging task than in generic domains [23].

In this paper, we address a relatively novel problem: the identification of cause-effect relationships and their arguments in the biomedical domain for Leukemia. Leukemia is a group of cancers that begins in the bone marrow and results in high number of abnormal white blood cells (WBC), called *leukemia cells*. Leukemia is the most frequent type of cancer in children. In 2015, Leukemia was detected in 2.3 million people and resulted in $353,500$ deaths; the average five-year survival rate is 57% in the USA.[4] The exact causes for Leukemia are unknown, although some risk factors are known, including family history, smoking, and exposure to ionizing radiation or chemicals such as *benzene*. Table 1 shows some example sentences about Leukemia in which CE relation mentions, *viz.*, Cause (C), Effect (E), and Causal-Cue (CC) are present. Note that sometimes the CE relation mention does not include a causative verb, but a causal *cue phrase*, such as *due to, because, hence, therefore*.

The CE relations in leukemia are at widely different abstraction levels - from genetic, molecular, cellular, organ level, tissue level to patient-level as an entity. In the corpus, we can discern a finer structure to the CE relations, apart from the two standard arguments, cause and effect. For example, CE relations seem have associated with them additional optional information, such as *evidence* (see Table 1 (2)), or a *control condition i.e.*, a condition under which the causal relation holds (see Table 1 (3)). In the biomedical domain, a cause is often an *agent* (such as an organism, drug, compound), an *event*, an *action* or a *condition*. An *event* is any change in the physical state or property of one or more named entity instances. A *condition* is broadly any property or state of one or more named entity instances, which is sustained over reasonably long periods. An effect is often an event or a condition.

In this paper, we conceptualize that in a causal sentence, *cause* temporally precedes the *effect*. Long Short Term Memory (LSTM) network is a deep neural network having recurrent connections between the layers. It is tailored to process the text having a temporal order of words. Therefore the temporal order of CE relations can be captured well by the Long Short Term Memory network, which makes it a potential technique for the identification of CE relations in the biomedical domain. We present a CE relation identification system for the

[4] https://en.wikipedia.org/wiki/Leukemia

(1) [Human T-cell leukemia virus type 1 (HTLV-1)]$_C$ [causes]$_{CC}$ a highly lethal [blood cancer]$_E$ or a chronic debilitating [disease of the spinal cord]$_E$.
(1) The [co-expression of p96 (ABL/BCR)]$_C$ [enhanced]$_{CC}$ the [kinase activity]$_E$ and as a consequence, the [transformation potential of p185 (BCR/ABL)]$_E$.
(3) While survival rates for ALL have improved, [central nervous system (CNS) relapse]$_C$ remains a significant [cause]$_{CC}$ of [treatment failure]$_E$ and [treatment-related morbidity]$_E$.
(2) Using both [pharmacologic and genetic assays]$_E$, we show here that [inactivation of RIP1/RIP3]$_C$ [resulted]$_{CC}$ in [reduction of SOCS1 protein levels]$_E$ and [partial differentiation of AML cells]$_E$.
(3) [Bone mass acquisition]$_E$ may be [compromised]$_{CC}$ in survivors of childhood acute lymphocytic leukemia due to various factors, including [adiposity]$_C$.
(6) [cCMP-AM]$_C$ did not [induce]$_{CC}$ [apoptosis in K-562 cells, a human chronic myelogenous leukemia cell line,]$_E$ [due to]$_{CC}$ [rapid export via multidrug resistance-associated proteins]$_C$.

Table 1. Examples of CE relations in leukemia.

biomedical domain with the focus on Leukemia. First the system is formalized as a binary classification system with two classes, *viz., CE-Relation, Not-CE-Relation*. Next the sentences which are identified as CE Relation tag are used for the extraction of CE Relations arguments using a rule based system. Figure 1 shows the architecture of the proposed system. Stage-1 is the neural binary classification model, which identifies whether a sentence has CE relation or not. Stage-2 performs extraction of CE relation constructs using a rule based system.

Though deep neural networks require a massive amount of labeled datasets for classification, the tagged data requirement is overcome by getting focused embeddings of words trained on a large unlabeled corpus specific to Leukemia. We compare our LSTM-based model with Multi-layer Perceptron (MLP) and Support Vector Machine (SVM) for CE relation identification using three types of word embeddings, *viz., GloVe, PubMed*, and *target-specific* where the target is Leukemia. Results (5) show that LSTM with target-specific embeddings outperformed all other reported models. Furthermore, we show that an ensemble of LSTM models trained using GloVe and target-specific embeddings gives a significant improvement ($\sim 3\%$) over the individual models.

The major contributions of the paper are as follows.

- We generate $2,01,066$ embeddings specific to Leukemia using $60,000$ research papers on Leukemia from PubMed. We show the effectiveness of these focused (target-specific) embeddings over pre-trained embeddings for CE relation identification task.
- An ensemble of LSTMs trained using GloVe and target-specific embeddings produces an accuracy of **83.78**%, which is significantly greater over the individual models for the CE relation identification task.
- We generate a knowledge-base of 277478 CE relation mentions from the dataset of $60,000$ documents.

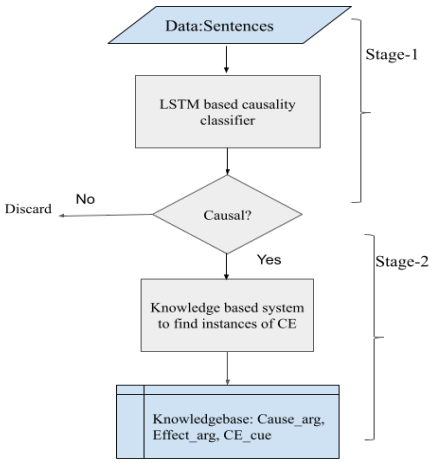

Fig. 1. Flowchart of the proposed system

The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 describes the preparation of training data and the semantic embeddings used in the paper. Section 4 provides the experimental setup. Section 5 shows the results and Section 6 concludes the paper.

2 Related Work

CE relation identification, in general, has been and continues to be well studied in the literature. Much of the work has attempted to discover CE relationships in the text by focusing on lexical and semantic constructs.

Kaplan et al., [11] wrote hand-coded rules considering causal scenario may vary from context to context. Joskowicz et al., [10] prepared a dedicated knowledge-base to build a causal analyzer for a Navy ship. Their objective was to understand a short narrative message about the Navy ship's equipment using CE relation. However, knowledge-based systems have low generalizability. In addition, building and maintaining a knowledge-based system is expensive for the targeted domain itself. Many researchers have used linguistic patterns to identify CE relations in the text without using any knowledge-base [7, 12]. A few works used grammatical patterns to identify CE relations [8, 22, 13]. There are very few instances of combining grammatical patterns with machine learning to extract semantic relations, such as cause-effect [6, 3]. In another work, cue phrases (cause

triggering construct) with their probability were used to extract other lexical arguments of cause-effect relation [3]. Do [6] developed a minimally supervised approach based on focused distributional similarity and discourse connectives.

None of the work discussed so far has considered the complications of the biomedical domain. However, due to domain-specific vocabulary and constructs, conventional CE relation extraction methods are not suitable in the biomedical domain. Mihuailua et al., [16] defined an annotation scheme for enriching biomedical domain corpora with causality relations. Their scheme was used to annotate 851 causal relations to form BioCause, a collection of 19 open-access full-text biomedical journal articles. Mihuailua et al., [15] created several baselines and experimented with and compare various parameter settings for three algorithms, *i.e.*, Conditional Random Fields (CRF), Support Vector Machines (SVM) and Random Forests (RF) for causality detection in the biomedical domain. They also evaluated the impact of lexical, syntactic, and semantic features on each of the algorithms, and showed that semantics improves the performance in all cases. Sharma et al., [23] proposed an approach that deploys the linguistic cue indicating CE constructs and PMI between dependency relations for identification of CE relation in a sentence.

Knowledge-based and pattern-based approaches have severe coverage issues. They can only consider those instances for which knowledge or pattern can be derived by observing the training data. This paper presents a deep-neural-network-based supervised approach, that is, LSTM for CE relation identification with target-specific word embeddings as input. The use of target-specific semantic embeddings of words facilitates capturing complex CE relations while reducing the need for excessive labeled data requirements.

3 Training Data and Embeddings

Leukemia is a highly researched disease in the biomedical domain, having more than $3,02,926$ scientific documents on PubMed and more than $3,09,492$ on Nature. As CE relations can be expressed using various semantic constructs, we use distributed representation of a sentence capturing various characteristics of the text in terms of embeddings and then use them for training classification models. The training dataset and the embeddings used in the paper are described below.

3.1 Training Data

We extracted a set of 2500 sentences from Leukemia-related papers in PubMed and labeled them for the training of models. Two competent annotators were consulted to assign binary labels: *CE Relation (1)* and *Not CE Relation (0)*. The *Cohens k* between the annotators is 0.97 [2]. We used majority voting to determine the actual label. Section 5 reports the 5-fold cross-validation accuracy on this dataset.

3.2 Generic GloVe Embeddings

The Global Vector model [21] referred to as GloVe combines word2vec with the ideas drawn from matrix factorization methods, such as LSA [5]. We used *pre-trained* GloVe word embeddings of size 300. We refer to them as *generic* embeddings as they are trained on the Wikipedia 2014 dataset.[5]

3.3 Target-specific Word Embeddings

To obtain target-specific word embeddings where the target (focus) is Leukemia, we parsed and downloaded 60,000 abstracts containing the term *Leukemia* from PubMed using the *Entrez* package of Biopython. The corpus has 5,12,061 sentences and 1,22,29,561 words. Embeddings of size 300 are learned from the corpus using the word2vec package [17]. Default parameters were used to train the model. The same dataset of 60,000 documents is used to prepare the knowledge-base containing CE relation arguments.

3.4 Domain-specific PubMed Embeddings

This is a set of pre-trained embeddings in the biomedical domain. The embeddings are trained on abstracts from PubMed without focusing on any particular disease.[6] Essentially, these semantic embeddings are trained using a domain-specific corpus, that is, the biomedical domain, but not specifically dedicated to the target for which classifier has to be trained, unlike our *target-specific word embeddings*.

Embeddings	Vocab-Size	Words-Found
Generic-GloVe	400000	4995
Domain-specific	1999860	6323
Target-specific	201066	6678
Training Data	6740	6740

Table 2. Statistics for the Word Embeddings

In addition to the above-described embeddings, we have observed the performance of LSTM with embeddings learned from the training dataset by LSTM's embeddings layer. Table 2 shows the statistics related to the embeddings used in this paper. Column 3 of Table 2 presents the number of words from the training dataset (Section 3.1) whose embeddings are found in the embeddings set.

[5] Download : https://nlp.stanford.edu/projects/glove/.
[6] Available for download: http://evexdb.org/pmresources/vec-space-models/.

4 Experimental Setup

This paper hypothesizes that the temporal order between cause-expression and effect-expression can be captured well by the Long Short Term Memory (LSTM) network. While Support Vector Machine (SVM) [15] and Multilayer Perceptron (MLP) are not tailored to process the sequential structure of words, hence not much suitable for CE relation identification. CE relation is a contextual property; LSTM generates hidden features representing the context. Hence it is a favorable architecture for CE relation identification. Lilleberg et al., [14] validated that word embeddings bring extra semantic features that help in text classification. Therefore, the use of independently trained semantic embeddings of words in place of words overcomes the labeled data requirement of LSTM. We provide a comparison among LSTM, SVM, and MLP using three types of semantic embeddings, *viz.*, GloVe, PubMed, and target-specific where the target is Leukemia.

To train an SVM based classifier, we have used the publicly available Python-based Scikit-learn package [20]. Though results are reported with linear kernels due to their superior performance, we experimented with other polynomial kernels. Yin and Jin, [24] speculated that the sum of word embeddings is meaningful and can represent the document. For example, the sum of word embeddings of *Germany* and *capital* is close to the embedding of *Berlin* [18]. We adhered to the same convention to produce embeddings of sentences to train SVM-based classifiers with embeddings (Equation 1).

$$\begin{aligned} &\text{Sentence (S) is having } t_i \text{ token with } v_i \text{ embedding:} \\ &S(t_1 : v_1; t_2 : v_2; ...; t_n : v_n), \\ &v_i \text{ is an } m \text{ dimension vector:} \\ &v_i = (v_{i1}, v_{i2}, ..., v_{im}), \\ &\text{Sentence embedding S of } m \text{ dimension:} \\ &S = (\sum_{i=1}^{n} v_{i1}, \sum_{i=1}^{n} v_{i2}, ..., \sum_{i=1}^{n} v_{im}) \end{aligned} \tag{1}$$

To implement MLP and LSTM, we used Keras functional API. The embedding layer of the LSTM network is initialized with the size of the embeddings. The middle layer is an LSTM layer, which is initialized with 256 activation units. The output layer is a dense layer having *sigmoid* as the activation function. MLP has the same settings, except the middle layer is a dense layer with 256 activation units.

Knowledge-base Generation: The instances classified as CE relation by our system become the input to a rule-based system for extraction of CE relation mentions. We use the rule-based system proposed by Sharma et al., [23] for this purpose. It is specifically trained in an unsupervised manner to extract CE relation mentions from the bio-medical text. It is based on the principle that a known causal verb can be used to extract CE arguments, and known CE

	t-value	P-value
LSTM-GloVe *vs* SVM-BoW	1.88	0.04
LSTM-Target-specific *vs* SVM-BoW	2.18	0.0
Ensemble *vs* SVM-BoW	5.39	0.00
Ensemble *vs* LSTM-Generic-GloVe	5.19	0.00
Ensemble *vs* LSTM-Target-specific	4.08	0.00

Table 3. t-test ($\alpha = 0.05$) results for the systems having significant difference in accuracy.

arguments can be used to discover unknown causative verbs (hence *co-discovery*). Point-wise mutual information (PMI) is used to measure the level of (linguistic) associations between a causative verb and its argument.

5 Results

We implemented 12 Systems to validate our hypothesis extensively. Figure 2 shows the 5-fold cross-validation accuracy concerning each system. **BoW** is the Bag-of-words model with SVM. **Train-MLP** and **Train-LSTM** are models trained on embeddings obtained from training data only with MLP and LSTM settings, respectively. Out of the remaining nine, three systems employ SVM, three use MLP, and three use LSTM, where each system in the collections individually trained using **PubMed** (Section 3.4), **GloVe** (Section 3.2), and **target-specific embeddings** (Section 3.3), respectively.

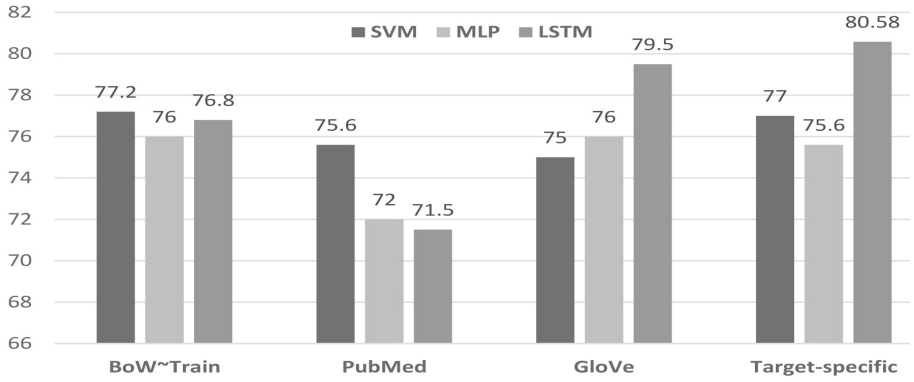

Fig. 2. 5-fold cross-validation accuracy in % for CE relation identification.

Mihuailua et al., [15] used Bag-of-words with SVM for causality identification. Figure 2 shows that SVM-BoW produces a 5-fold cross-validation accuracy of 77.2%. On the other hand, the performance of SVM with PubMed, GloVe, and target-specific embeddings is not significantly different from that of SVM-BoW. SVM is not able to incorporate the additional semantic information and contextual information provided by embeddings. The difference in the vocabulary of diseases (*eg.*, Leukemia and Glioma have many dedicated words) makes PubMed embeddings (Section 3.4) inadequate for finding CE relations in Leukemia. On the other hand, LSTM with GloVe outperforms SVM-BoW by a significant margin. These generic word embeddings bring in additional favorable information that is not available in the training data. On the other hand, target-specific embeddings that are obtained from the data focusing on the targeted disease (Leukemia) performed the best with LSTM.

The performance of LSTM-Train is inferior to that of SVM-BoW as the data is not sufficiently large for LSTM. Use of pre-trained embeddings *viz.*, generic-GloVe, and target-specific embeddings reduce the labeled data requirement of LSTM. On the other hand, the performance of MLP is not significantly different from SVM. Both the algorithms are unable to capture the context formed by the sequence of words. CE relation has a long term dependency, *cause*, *effect* and *causality cue* mentions can be any words apart in the sentence. Long Short Term Memory (LSTM) network solves this problem by using gates to control the memorizing process [9].

Ensemble: We observed that an ensemble of LSTM models trained using generic *GloVe* and *target-specific* embeddings produced an accuracy of **83.78%**, which is significantly greater than the accuracy delivered by the individual classifier for CE relation identification. The classification probability value assigned by the individual classifier is averaged to obtain the ensemble classification probability. If the averaged probability is more than 0.5 for any instance, it is classified as having CE relation (1), else not having CE relation (0). Essentially, both the systems bring in complementary information as their embeddings are trained on two completely different corpora; the first is generic (Wikipedia) corpus, another is Leukemia (target) corpus. Table 3 shows the t-test results for pairs of systems where the first system performs significantly better than the second system. LSTM with generic-GloVe and target-specific embeddings are observed to be significantly better than any other system, including SVM-BoW as per t-test. LSTM-Golve and LSTM-Target-specific models' ensemble reported a significant improvement over individual models as per t-test.

Table 4 presents the statistics related to the knowledge-base obtained from the instances classified as CE relation by our LSTM-based ensemble system. CE relation mentions which are forming the knowledge-base are identified using using the approach proposed by Sharma et al., [23].

footnotesize

$\mathbf{CE}_{Cue}$	$\mathbf{CE}_{Cause}$	$\mathbf{CE}_{Effect}$
98778	87235	91465

Table 4. Statistics of Knowledge-base

6 Conclusion

Cause-Effect (CE) relation in a scientific text is an instance of knowledge required to be identified to answer causal questions. In this paper, we present that the long-term dependency between *cause* and *effect* expressions in a sentence can be captured well by the LSTM network for CE relation identification. The use of target-specific embeddings, which are learned from a corpus focused on the targeted disease, overcomes the labeled data requirement of LSTM. In addition, embeddings learned from a generic corpus (Wikipedia), *i.e.*, GloVe provides complementary information to the model. Results show that LSTM with target-specific embeddings and GloVe produce 80.5% and 79.5% accuracy, respectively, which is significantly better than models trained using Support Vector Machine and Multilayer Perceptron. Furthermore, an ensemble of the LSTM models trained using GloVe and target-specific embeddings produced an accuracy of 83.7%, which is significantly greater than the accuracy delivered by the individual classifier for CE relation identification. Furthermore, our CE relation classification system's results generate a knowledge-base of 277478 CE relation mentions using a rule-based approach.

References

1. Ananiadou, S., Mcnaught, J.: Text mining for biology and biomedicine. Citeseer (2006)
2. Berry, K.J., Mielke Jr, P.W.: A generalization of cohen's kappa agreement measure to interval measurement and multiple raters. Educational and Psychological Measurement **48**(4), 921–933 (1988)
3. Chang, D.S., Choi, K.S.: Causal relation extraction using cue phrase and lexical pair probabilities. In: International Conference on Natural Language Processing. pp. 61–70. Springer (2004)
4. Cohen, K.B., Hunter, L.: Getting started in text mining. PLoS computational biology **4**(1) (2008)
5. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American society for information science **41**(6), 391 (1990)
6. Do, Q.X., Chan, Y.S., Roth, D.: Minimally supervised event causality identification. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 294–303. Association for Computational Linguistics (2011)
7. Garcia, D., et al.: Coatis, an nlp system to locate expressions of actions connected by causality links. In: International Conference on Knowledge Engineering and Knowledge Management. pp. 347–352. Springer (1997)

8. Girju, R.: Automatic detection of causal relations for question answering. In: Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering-Volume 12. pp. 76–83. Association for Computational Linguistics (2003)
9. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
10. Joskowicz, L., Ksiezyck, T., Grishman, R.: Deep domain models for discourse analysis. In: AI Systems in Government Conference, 1989., Proceedings of the Annual. pp. 195–200. IEEE (1989)
11. Kaplan, R.M., Berry-Rogghe, G.: Knowledge-based acquisition of causal relationships in text. Knowledge Acquisition **3**(3), 317–337 (1991)
12. Khoo, C.S., Kornfilt, J., Oddy, R.N., Myaeng, S.H.: Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. Literary and Linguistic Computing **13**(4), 177–186 (1998)
13. Kim, H.D., Zhai, C., Rietz, T.A., Diermeier, D., Hsu, M., Castellanos, M., Ceja Limon, C.A.: Incatomi: integrative causal topic miner between textual and non-textual time series data. In: Proceedings of the 21st ACM international conference on Information and knowledge management. pp. 2689–2691. ACM (2012)
14. Lilleberg, J., Zhu, Y., Zhang, Y.: Support vector machines and word2vec for text classification with semantic features. In: Cognitive Informatics & Cognitive Computing (ICCI* CC), 2015 IEEE 14th International Conference on. pp. 136–140. IEEE (2015)
15. MIHĂILĂ, C., Ananiadou, S.: Recognising discourse causality triggers in the biomedical domain. Journal of bioinformatics and computational biology **11**(06), 1343008 (2013)
16. Mihăilă, C., Ohta, T., Pyysalo, S., Ananiadou, S.: Biocause: Annotating and analysing causality in the biomedical domain. BMC bioinformatics **14**(1), 2 (2013)
17. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
18. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
19. Moldovan, D., Paşca, M., Harabagiu, S., Surdeanu, M.: Performance issues and error analysis in an open-domain question answering system. ACM Transactions on Information Systems (TOIS) **21**(2), 133–154 (2003)
20. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**(Oct), 2825–2830 (2011)
21. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: EMNLP. vol. 14, pp. 1532–43 (2014)
22. Radinsky, K., Davidovich, S., Markovitch, S.: Learning causality from textual data. In: Proceedings of Learning by Reading for Intelligent Question Answering Conference (2011)
23. Sharma, R., Palshikar, G., Pawar, S.: An unsupervised approach for cause-effect relation extraction from biomedical text. In: International Conference on Applications of Natural Language to Information Systems. pp. 419–427. Springer (2018)
24. Yin, Y., Jin, Z.: Document sentiment classification based on the word embedding. In: 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering (2015)