Automatic Cause-Effect Relation Extraction from Dental Textbooks Using BERT

Terapat Chansai[1], Ruksit Rojpaisarnkit[1], Teerakarn Boriboonsub[1], Suppawong Tuarob[1(✉)], Myat Su Yin[1], Peter Haddawy[1,4], Saeed-Ul Hassan[2], and Mihai Pomarlan[3]

[1] Faculty of Information and Communication Technology, Mahidol University, Nakhon Pathom, Thailand
{terapat.chn,ruksit.roj,teerakarn.bor}@student.mahidol.edu, {suppawong.tua,myatsu.yin,peter.had}@mahidol.ac.th

[2] Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, UK
S.Ul-Hassan@mmu.ac.uk

[3] Faculty of Linguistics, University of Bremen, Bremen, Germany
mihai.pomarlan@uni-bremen.de

[4] Bremen Spatial Cognition Center, University of Bremen, Bremen, Germany

Abstract. The ability to automatically identify causal relations from surgical textbooks could prove helpful in the automatic construction of ontologies for dentistry and building learning-assistant tools for dental students where questions about essential concepts can be auto-generated from the extracted ontologies. In this paper, we propose a neural network architecture to extract cause-effect relations from dental surgery textbooks. The architecture uses a transformer to capture complex causal sentences, specific semantics, and large-scale ontologies and solve sequence-to-sequence tasks while preserving long-range dependencies. Furthermore, we have also used BERT to learn word contextual relations. During pre-training, BERT is trained on enormous corpora of unannotated text on the web. These pre-trained models can be fine-tuned on custom tasks with specific datasets. We first detect sentences that contain cause-effect relations. Then, cause and effect clauses from each cause-effect sentence are identified and extracted. Both automatic and expert-rated evaluations are used to validate the efficacy of our proposed models. Finally, we discuss a prototype system that helps dental students learn important concepts from dental surgery textbooks, along with our future research directions.

Keywords: Relation extraction · Textbook mining · Deep learning

1 Introduction

Intelligent tutoring systems have been applied to a wide variety of domains, including clinical medicine. A particular challenge in medical applications is acquiring

H.-R. Ke et al. (Eds.): ICADL 2021, LNCS 13133, pp. 127–138, 2021.
https://doi.org/10.1007/978-3-030-91669-5_11

the structured knowledge needed to teach the required concepts. The teaching of clinical and surgical procedures requires providing students with a thorough understanding of causal relations between actions and their possible effects in the context of various states of the patient so that students can generalize beyond the particular scenario being presented. Such knowledge is available in fairly structured form in medical textbooks. If it could be automatically extracted and suitably structured, this would address a major bottleneck in building intelligent tutoring systems for clinical medicine and surgery. Yet, identifying causal relations in textbooks is challenging since causality can be linguistically expressed in many, sometimes ambiguous and wordy, ways. Furthermore, building a precise model to automatically extract causal relations requires good quality datasets and experts to analyze results concerning a large variety of sentences. For example, in the sentence "if undetected, transportation may lead to ledging, zipping, gouging, or even perforation of the root canal wall", the event "transportation" causes the events "ledging, zipping, gouging, or even perforation of the root canal wall". In a more complicated scenario, "gouging the root canal wall" can also be the effect of "forcing rotary instruments such as Gates-Glidden burs beyond the resistance level", therefore there are several potential causes for an effect, as well as more complex chains of causation, which might lead to the entity linking problem which will be discussed in a later section. The sentences above are from one endodontics textbook commonly used for teaching dental students.

Many approaches have been employed to extract machine-readable information from natural language text. Some of these used rule-based extraction models, while some used traditional machine learning algorithms. Regardless, most of the existing methods expect the participants to be denoted by one name each in one sentence expressing the relationship, limiting their applicability to natural text. Novel techniques like deep learning have only begun to be applied to identify and extract causal relations.

In this work, we address the problem of extracting causal relations from textbooks on dental endodontic procedures. We choose endodontics because it is one of the more challenging areas of dentistry and thus more difficult to learn. Our approach combines recent strategies to deal with NLP problems. Specifically, we employ BERT [4] and Transformers [18] to address the causal relation extraction problem. We propose a neural network architecture to capture such complex sentences in dental surgery textbooks and use BERT to learn contextual causal relations by training it with a rich dataset. We identify two subproblems in this task, cause-effect sentence classification (recognize when a sentence expresses a causal relation) and cause-effect relation extraction (once the relation is identified, identify its participants). Both of these are discussed in Sect. 3. The source code is made available for research purposes at: https://gitlab.com/dentoi/casual-extraction. Although our application domain in this paper is endodontics, the presented approach is quite general and applicable to other clinical and surgical domains.

2 Background and Related Work

While machine learning techniques have been extensively used in text mining to extract meaningful information from large-scale corpora [14,15], we briefly describe concepts and techniques from previous research that we use in this work. An *ontology* represents a formalization of a conceptualization, i.e., a description in some formal language of entities and relations between them. It is used for knowledge representation for discrete intelligent reasoning and to exchange such knowledge across systems that may use it. *Information extraction* is a task for extracting knowledge from unstructured text data, then converting it into a structured, machine-readable representation such as a semantic graph. The process consists of several steps such as Named Entity Recognition, Named Entity Linking, and Relation Extraction. *BERT (Bidirectional Encoder Representation from Transformer)* is a recent but prevalent technique in NLP to implement a language model, i.e., a model capable of assessing how likely a particular text is. Furthermore, BERT has proven to be applicable for a whole range of NLP tasks, including question answering and relation extraction. BERT is based on the Transformer model architecture, which uses attention mechanisms that learn contextual relationships between each token in a text. A transformer consists of an encoder to read the text input and a decoder to produce a prediction for the task.

Previous work has used several approaches to extract causal relations from the text in a variety of domains. Dasgupta et al. [3] used a recursive neural network architecture (Bi-directional LSTM) with word-level embeddings to detect causal relations in a sentence. A conditional random field (CRF) model was used to evaluate performance. They used a diverse trained dataset, including drug effect, BBC News Article, and SemEval.

Su Yin et al. [13] explored the possibility of automatically extracting causal relations from textbooks on dental procedures. They used pattern-matching over dependency parse trees produced by the spaCy NLP tool to identify causal relation assertions in a collection of fifteen textbooks on endodontic root canal treatment. Since their primary purpose was to extract knowledge for teaching, they focused on surgical mishaps and their possible causes. They achieved a precision of 95.7% but a recall of only 41.6%.

Zhao et al. [20] sought to extract causal transitive relations in medical text data, where one effect is a cause in another relation. First, they examined how such chains of causal relations are expressed over several sentences. Their approach used causal triads, which are defined as associations of co-occurring medical entities that are likely to be causality connected. Then, the authors tried to discover implicit causal relations. The result showed that the causal triad model is an appropriate means to extract textual medical causal relations.

Hybrid techniques have also been used to automatically extract causal relations from text [12]. The hybrid system combines both rule-based and machine learning methodologies; it extracts cause-effect pairs utilizing a set of rules for Simple Causative Verbs, rules for Phrasal Verbs/Noun, etc. Using the dataset SemEval-2010 (Task 8), they obtained 87% precision and 87% recall. Therefore, the rule-based model can also be adapted for other types of information, such as the dental textbooks that we use in this work.

3 Methodology

Our proposed methodology is composed of three modules: a) data preprocessing, b) cause-effect sentence classification, and c) cause-effect relation extraction. Each of the modules is described in the following sub-sections.

3.1 Dataset and Data Preprocessing

We used 16 dental surgery textbooks as our dataset to train the model. We first converted the PDF file to a text file and cleaned non-ASCII characters out of the text file for our data preparation processes. We then split each text file into several chapters using *chapterize*[1]. As a result, each chapter's text file was produced, consisting of several paragraphs and sentences, ready to use for model training. The dataset contains 5,642 annotated sentences, of which 2,032 sentences express a cause-effect relation. The sentences describe endodontic procedures and are labeled using the pattern-based method proposed by Su Yin, et al. [13], shown to yield high precision of 95.7%, but low recall of 41.57%.

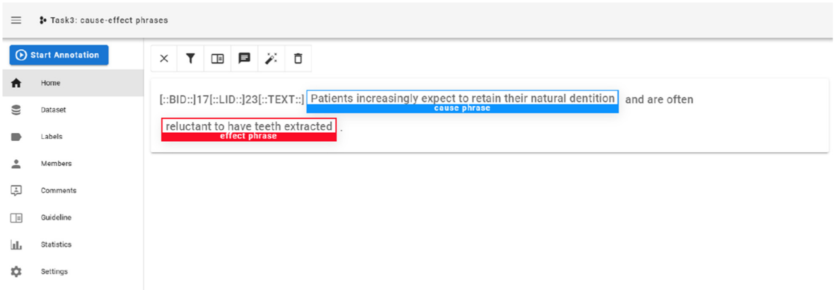

Fig. 1. Annotation tool for the cause-effect relation extraction task.

For each sentence that expresses a cause-effect relation, human annotators were asked to label parts of the sentence (word sequences) which represent cause and effect clauses. Doccano[2], an open source document annotation framework, was used to facilitate the annotation, as presented in Fig. 1.

3.2 Cause-Effect Sentence Classification

Cause-effect sentence classification is the method to identify whether or not a sentence expresses a cause-effect relation. Each sentence in the textbook is encoded before being used to train the BERT model. For the implementation of the cause-effect sentence classification model, there are four architectures. First, the input encoder consists of input_ids, input_mask, and segment_ids. Second, the BERT model consists of 24 layers and 1,024 hidden units in each layer.

[1] https://github.com/JonathanReeve/chapterize.
[2] https://github.com/doccano.

Third, a linear layer for the sigmoid activation function is used for formatting BERT's output then passed to the classifier, the last architecture in this model. The classifier uses the sigmoid activation function for binary classification. The 5,642 sentences used to develop this model are separated into 70% for training, 10% for validation, and 20% for testing.

Figure 2 illustrates the pipeline for model training and classification process, along with the model's optimal hyperparameters. First the annotated training data is used to train a BERT model, which is then fine-tuned with the validation dataset. The trained classifier takes a sentence as the input and outputs the probability of being a causal sentence.

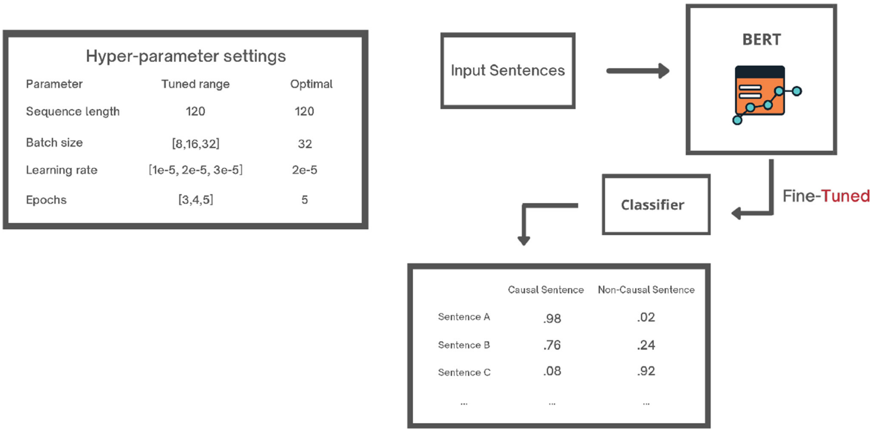

Fig. 2. The processing pipeline for the sentence classification task.

As baselines to BERT, we have also tried other deep learning based (i.e., CNN for sentence classification [7]) and traditional machine learning classifiers such as Naive Bayes [19], Support Vector Machine (SVM) [8], and Random Forest [1] with default hyperparameter settings. For traditional machine learning methods, each sentence is represented with a bag-of-word vector using TF-IDF weights.

3.3 Cause-Effect Relation Extraction

Cause-Effect relation extraction is the method to extract the cause and the effect from sentences expressing causal relations in textbooks. The causal sentence is tokenized into three types; C stands for Cause, E stands for Effect, and O stands for Other. Each token is labeled and encoded before training in the BERT model. For the implementation of the Cause-Effect Relation Extraction model, there are four architectures. The first and second architectures are the same as in the Cause-Effect Sentence Classification model. The third is the dropout layer, which differs from the Cause-Effect Sentence Classification model because the binary classification model is valid only for the first index of BERT's output. The last architecture is the classifier, which consists of three classes: cause, effect, and other, as mentioned earlier. The dataset for this model is the same that for the

Cause-Effect Sentence Classification model. As a baseline, we also compare the performance of the BERT model with a Bi-LSTM CRF model proposed in [10], due to its reported effectiveness in sequence tagging tasks such as named entity recognition [6].

Figure 3 illustrates the pipeline for model training and extraction process along with the model's optimal hyperparameters. We framed the cause-effect relation extraction task as a token classification problem, where each token is classified into Cause, Effect, or Other. Therefore, the BERT model has been configured for the multiclass classification task.

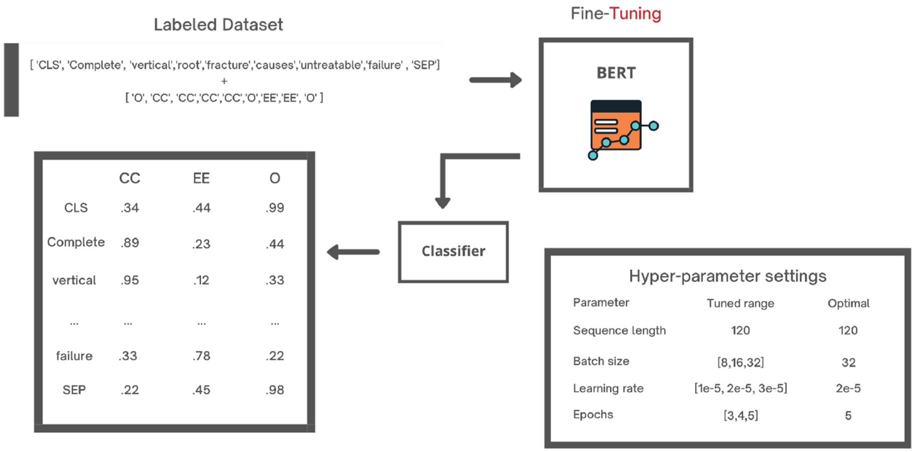

Fig. 3. The processing pipeline for the cause-effect extraction task.

4 Experiments, Results, and Discussion

All the experiments were conducted on a Linux machine with 16 CPU cores (32 threads), two RTX 2080 Super GPUs, and 128 GB of RAM. This section discusses the results.

4.1 Cause-Effect Sentence Classification

In cause-effect classification, each sentence is labeled as 0 or 1 for Non-causal or Causal sentence, respectively. Table 1 shows examples of causal and non-causal sentences taken from an endodontics textbook.

Table 1. Examples of causal and non-causal sentences, taken from [11].

Sentence	Label
Repeated recapitulation or remaining stationary in the canal with a non-landed rotary instrument can lead to apical transportation	1
A deep palate allows much greater vertical access when using a palatal approach	0
Furthermore, this pain can be further exacerbated by incorrect or unnecessary treatments, often resulting in the establishment of chronic pain pathways	1
This can be started 24 h prior to surgery but can also be swilled for 1 min prior to placement of anaesthetic	0

The precision, recall, and F1 scores for each class (Causal, Non-Causal, Weighted Average) predicted by BERT and other classifiers are reported in Table 2. It is apparent that deep learning based methods such as CNN and BERT outperform the traditional machine learning methods (i.e., Naïve Bayes, SVM, and Random Forest) in all aspects. While the CNN has a slightly better recall for the Causal Sentence class compared to BERT, BERT yields the highest F1 in both the Non-Causal and Causal Sentence classes. This may be due to BERT's ability to encode proximity semantics in a sentence into the embedding, making it suitable for tasks that require an understanding of word sequences and language syntax.

Table 2. Performance of the causal sentence classification task.

Model	Non causal sentence			Causal sentence			Weighted average		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Naive Bayes	0.81	0.98	0.89	0.86	0.31	0.45	0.82	0.78	0.82
SVM	0.95	0.94	0.95	0.84	0.86	0.85	0.93	0.93	0.93
Random Forest	0.95	0.94	0.95	0.84	0.86	0.85	0.93	0.92	0.92
CNN	**0.98**	0.95	0.96	0.83	**0.94**	0.88	0.95	0.95	0.95
BERT	0.97	**0.99**	**0.98**	**0.98**	0.90	**0.93**	**0.97**	**0.97**	**0.97**

4.2 Cause-Effect Relation Extraction

In the cause-effect relation extraction process, sentences are tokenized, and special tokens ("CLS" and "SEP") are added to the sentences. The "CLS" token annotates the sentences' starting point. The "SEP" token represents the ending of the sentences. The model's output is the vector representation of each sentence, which is later converted to each token label. The labels include 'C', 'E', and 'O', which represent Cause token, Effect token, and Other token, as illustrated in Table 3.

Table 3. Example labeling of cause and effect tokens.

Sentence	['CLS', 'Failure', 'to', 'achieve', 'patency', 'during', 'preparation', 'can', 'result', 'in', 'inadequate', 'penetration', 'of', 'irrigants', 'SEP']
Label	['O', 'C', 'C', 'C', 'C', 'C', 'C', 'O', 'O', 'O', 'E', 'E', 'E', 'E', 'O']

The precision, recall, and F1 of the cause-effect extraction task using Bi-LSTM CRF and BERT are reported in Table 4. It is evident that BERT outperforms Bi-LSTM CRF (baseline) in all metrics, both in the Cause Token and Effect Token classification tasks. Bi-LSTM CRF performs relatively well on extracting Cause tokens, but its performance drops when extracting the Effect tokens. Furthermore, Bi-LSTM CRF only yields F1 of 0.49 on identifying the Effect tokens. In contrast, BERT performs well on identifying both the Cause and Effect tokens, yielding an F1 of 0.89 on average.

Table 4. Performance of the cause-effect extraction task.

Model	Cause token			Effect token			Weighted average		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Bi-LSTM CRF	0.77	0.80	0.79	0.61	0.37	0.49	0.71	0.64	0.66
BERT	**0.89**	**0.91**	**0.90**	**0.89**	**0.90**	**0.89**	**0.89**	**0.89**	**0.89**

5 Future Directions

This paper has reported a primary investigation on extracting an ontology of causal relations from endodontics textbooks. Our overarching goal is to develop a toolbox capable of automatically extracting and structuring essential causal information from textbooks on clinical and surgical procedures. As future work, we plan to evaluate the result of the proposed approach by asking experts to assess our model's outcome regarding whether the extracted sentences contain a causal event or not. Subsequently, machine learning models could be further enhanced for textbooks in other languages using language-agnostic BERT-based models [5]. Besides improving the models, the subsections discuss our path forward in terms of implementation.

5.1 Relation Linking

Illustrated in Fig. 4 is a partial snapshot of extracted relations from individual causal sentences, visualized by *react-force-graph*[3]. Notable extracted relations include `post preparation` that leads to `root structures`, that further leads to `pulp therapy`. On the other branch, `access cavity` also leads to `pulp therapy` and `line access factors`. Linking these cause-effect relations together would give rise to many useful applications. Relation linking is a method to link each relevant entity to construct a knowledge graph, where several knowledge graph

[3] https://github.com/vasturiano/react-force-graph.

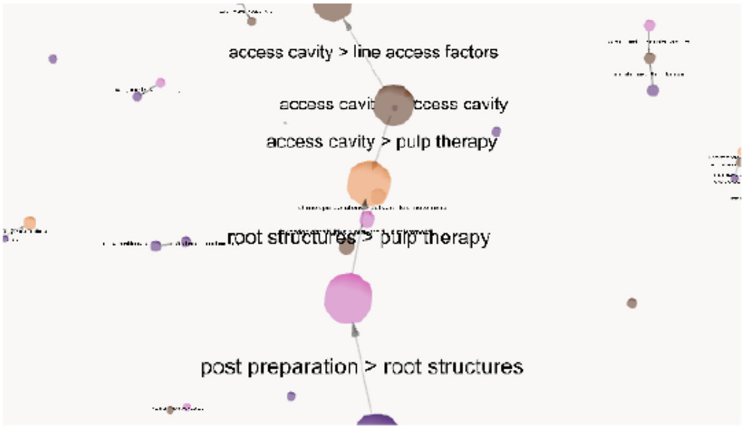

Fig. 4. Examples of extracted relations.

mining techniques can be applied to extract meaningful patterns. Our model might identify several cause-effect pairs that are similar or occur in a chain of causation. For example, if two cause-effect pairs have a similar cause, the effects of that two causes might be related. If we could link each relevant entity or cause together, it would be an efficient structure that effectively illustrates the knowledge and information. For example, linking causal relations and the conditions under which they hold with actions could enable reasoning about possible consequences of actions. Furthermore, it could be used for knowledge visualization or ontology visualization in a subsequent application.

5.2 Application in Dental Quiz Generation

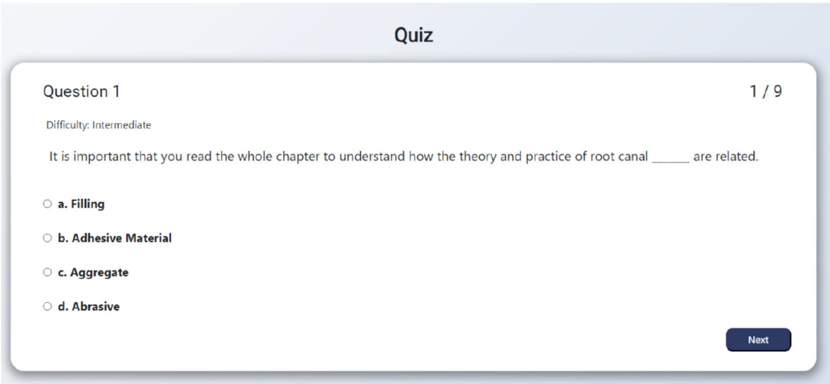

Fig. 5. Example of a dental quiz question.

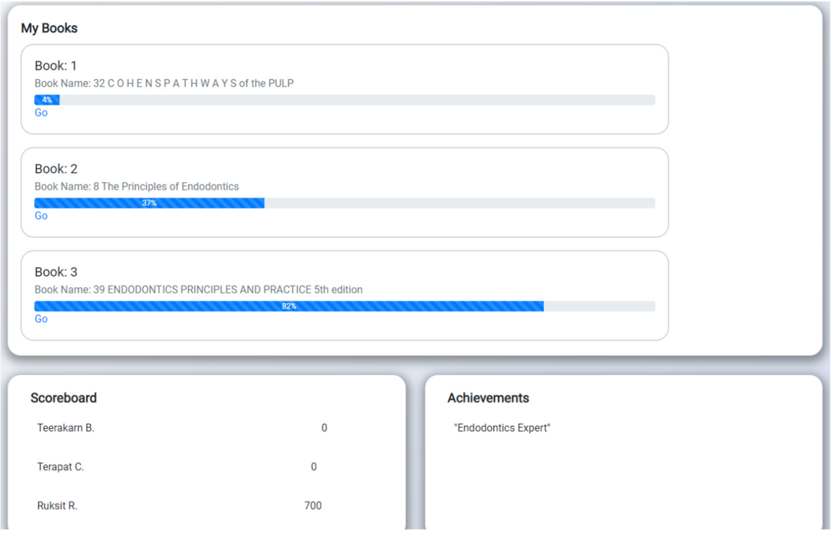

Fig. 6. Example dashboard showing a student' progress on completing quizzes from each textbook.

As dental students must learn from many textbooks, there are some difficulties distinguishing events and understanding information. Further, the method to assess the understanding of each topic is somewhat technical and requires an expert to construct the test. From the results of our proposed model, cause and effect pairs can be an efficient knowledge base that we could apply to generate a question and answering system, which could serve as a personal learning assistant to increase students' understanding and enhance their performance. Implementing a web-based application (similar to the work of Budovec et al. [2] in the field of radiology) could be an effective way for accessibility and quality. Quizzes can be generated from the cause and effect pairs that have been extracted from textbooks, which would be useful for dental students to assess their understanding. Automatic question-answer pair generation [9] could be adopted for this task. Figure 5 illustrates an example quiz question that could be helpful to students to assess their understanding of the textbook's material.

Dental instructors could also benefit from a system that allows them to automatically analyze each textbook and generate quizzes for each chapter. These quizzes can be assigned to students where the system can keep track of each student's progress on their self-study on each textbook, as illustrated in Fig. 6.

Recently Vannaprathip et al. [16,17] have presented an intelligent tutoring system for teaching surgical decision making, with application in the domain of endodontics. The intelligent tutor is integrated with a VR dental simulator. The tutor intervenes during the surgical procedure by pointing our errors, providing positive feedback, and asking a variety of types of questions. To do this, it makes use of causal knowledge represented in terms of conditional effects of actions using an adaptation of the PDDL AI planning language. The coding of this domain knowledge was a major bottleneck in building the automated tutoring system. The techniques presented in this paper represent a step toward automating that knowledge representation process.

6 Conclusions

This paper has proposed a neural network architecture to extract cause-effect relations from endodontics textbooks and built a precise learning-assistant tool for dental students. The ultimate goal is to teach students to make decisions in novel situations by providing them with a complete understanding of causal relations occurring in dental procedures, relations which are described in textbooks. This understanding can be assessed through quizzes that are automatically generated from the ontologies extracted by our approach.

Acknowledgments. This work was partially supported by a grant from the Mahidol University Office of International Relations to P. Haddawy in support of the MIRU joint unit. S. Tuarob is supported by Mahidol University (Grant No. MU-MiniRC02/2564). We also appreciate the computing resources from Grant No. RSA6280105, funded by Thailand Science Research and Innovation (TSRI) and the National Research Council of Thailand (NRCT).

References

1. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
2. Budovec, J.J., Lam, C.A., Kahn, C.E., Jr.: Informatics in radiology: radiology gamuts ontology: differential diagnosis for the semantic web. Radiographics **34**(1), 254–264 (2014)
3. Dasgupta, T., Saha, R., Dey, L., Naskar, A.: Automatic extraction of causal relations from text using linguistically informed deep neural networks. In: Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue, pp. 306–316 (2018)
4. 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)
5. Feng, F., Yang, Y., Cer, D., Arivazhagan, N., Wang, W.: Language-agnostic BERT sentence embedding. arXiv preprint arXiv:2007.01852 (2020)
6. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991 (2015)
7. Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746–1751. Association for Computational Linguistics, Doha, October 2014. https://doi.org/10.3115/v1/D14-1181. https://www.aclweb.org/anthology/D14-1181
8. Noble, W.S.: What is a support vector machine? Nat. Biotechnol. **24**(12), 1565–1567 (2006)
9. Noraset, T., Lowphansirikul, L., Tuarob, S.: WabiQA: a Wikipedia-based Thai question-answering system. Inf. Process. Manag. **58**(1), 102431 (2021)

10. Panchendrarajan, R., Amaresan, A.: Bidirectional lstm-crf for named entity recognition. In: PACLIC (2018)
11. Rhodes, J.S.: Advanced Endodontics: Clinical Retreatment and Surgery. CRC Press (2005)
12. Sorgente, A., Vettigli, G., Mele, F.: A hybrid approach for the automatic extraction of causal relations from text. In: Lai, C., Giuliani, A., Semeraro, G. (eds.) Emerging Ideas on Information Filtering and Retrieval. SCI, vol. 746, pp. 15–29. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-68392-8_2
13. Su Yin, M., et al.: Automated extraction of causal relations from text for teaching surgical concepts. In: Proceedings of 8th IEEE International Conference on Healthcare Informatics, November 2020
14. Tuarob, S., Mitrpanont, J.L.: Automatic discovery of abusive Thai language usages in social networks. In: Choemprayong, S., Crestani, F., Cunningham, S.J. (eds.) ICADL 2017. LNCS, vol. 10647, pp. 267–278. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70232-2_23
15. Tuarob, S., et al.: DAViS: a unified solution for data collection, analyzation, and visualization in real-time stock market prediction. Financ. Innov. **7**(1), 1–32 (2021)
16. Vannaprathip, N., Haddawy, P., Schultheis, H., Suebnukarn, S.: Intelligent tutoring for surgical decision making: a planning-based approach. Int. J. Artif. Intell. Educ. (2021, to appear)
17. Vannaprathip, N., et al.: A planning-based approach to generating tutorial dialog for teaching surgical decision making. In: Nkambou, R., Azevedo, R., Vassileva, J. (eds.) ITS 2018. LNCS, vol. 10858, pp. 386–391. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91464-0_44
18. Vaswani, A., et al.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
19. Zhang, H.: Exploring conditions for the optimality of naive bayes. Int. J. Pattern Recognit. Artif. Intell. **19**(02), 183–198 (2005)
20. Zhao, S., Jiang, M., Liu, M., Qin, B., Liu, T.: CausalTriad: toward pseudo causal relation discovery and hypotheses generation from medical text data. In: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, pp. 184–193 (2018)