Constructing and Interpreting Causal Knowledge Graphs from News

Fiona Anting Tan[1], Debdeep Paul[2], Sahim Yamaura[2], Miura Koji[2], See-Kiong Ng[1]
[1] Institute of Data Science National University of Singapore
[2] Panasonic Industrial Devices Singapore
tan.f@u.nus.edu, debdeep.paul@sg.panasonic.com

Abstract

Many financial jobs rely on news to learn about causal events in the past and present, to make informed decisions and predictions about the future. With the ever-increasing amount of news available online, there is a need to automate the extraction of causal events from unstructured texts. In this work, we propose a methodology to construct causal knowledge graphs (KGs) from news using two steps: (1) Extraction of Causal Relations, and (2) Argument Clustering and Representation into KG. We aim to build graphs that emphasize on recall, precision and interpretability. For extraction, although many earlier works already construct causal KGs from text, most adopt rudimentary pattern-based methods. We close this gap by using the latest BERT-based extraction models alongside pattern-based ones. As a result, we achieved a high recall, while still maintaining a high precision. For clustering, we utilized a topic modelling approach to cluster our arguments, so as to increase the connectivity of our graph. As a result, instead of 15,686 disconnected subgraphs, we were able to obtain 1 connected graph that enables users to infer more causal relationships from. Our final KG effectively captures and conveys causal relationships, validated through experiments, multiple use cases and user feedback.

Introduction

Many financial positions involve decision-making and prediction-making, like in stock price forecasting or sales and demand planning. Such careers require stakeholders to have a general understanding of the past, current and future market, and how one thing leads to another. Many stakeholders keep abreast of market trends by reading news. However, given the volume of the text online today, and even more if we were to consider historical news, it is impossible for any individual to consume all available information effectively.

In the past decade, knowledge graphs (KGs) have emerged as a useful way to store and represent knowledge. By performing end-to-end causal text mining (CTM) and then representing the causal relations through a KG, it is possible to summarize the past and current events succinctly for stakeholders to learn from effectively. We define end-to-end CTM as the identification of Cause and Effect arguments in any given text, if present.

In this paper, we focus on the application of summarizing and tracking causal relations in industry news to help individuals who frequently monitor the news for market research and decision making. Therefore, the final KG constructed must be useful by being: (1) **recall-focused:** it captures a large proportion of the causal relationships present in the news, (2) **precision-focused:** the causal relationships captured are truly causal, and (3) **interpretable:** it can be used by humans to learn causal relationships. Our methodology comprises of two broad steps, shown in Figure 1: (1) Extraction of Causal Relations, and (2) Argument Clustering and Representation into Knowledge Graph.

Our contributions are as follows:

- Although many earlier works investigate construction of causal KGs from text, most utilize pattern-based methods. In our work, we employ both pattern-based and neural network-based approaches. Our findings show that the pattern-based approach drastically misses out on extracting valid causal relations compared to the neural network-based approach (1:19 ratio).
- Graphs built directly off extracted Cause and Effect arguments are sparse and hence, hard to interpret. To mitigate this, we investigate a simple but effective solution to cluster our arguments based on semantics to create a more connected KG that enables more causal relationships to be drawn.
- We evaluate our methodology on a small set of data annotated by the users, demonstrate industry use cases and discuss users' feedback on the final KG. We intend to deploy our system as a regular service to the Sales Division.

The subsequent portions of the paper are outlined as follows: We first discuss related work. Subsequently, we introduce our data and describe our methodology. Next, we present our experimental results, demonstrate multiple use cases, and finally, conclude.

Related Work

In the recent years, many SOTA NLP solutions have been created to beat the leaderboards (Chen et al. 2022a; Zuo et al. 2020, 2021a,b; Cao et al. 2021; Chen et al. 2022b; Nik et al. 2022; Aziz, Hossain, and Chy 2022). Consistent with the general trend in NLP, the best models all use neural network architectures and pre-trained language models.

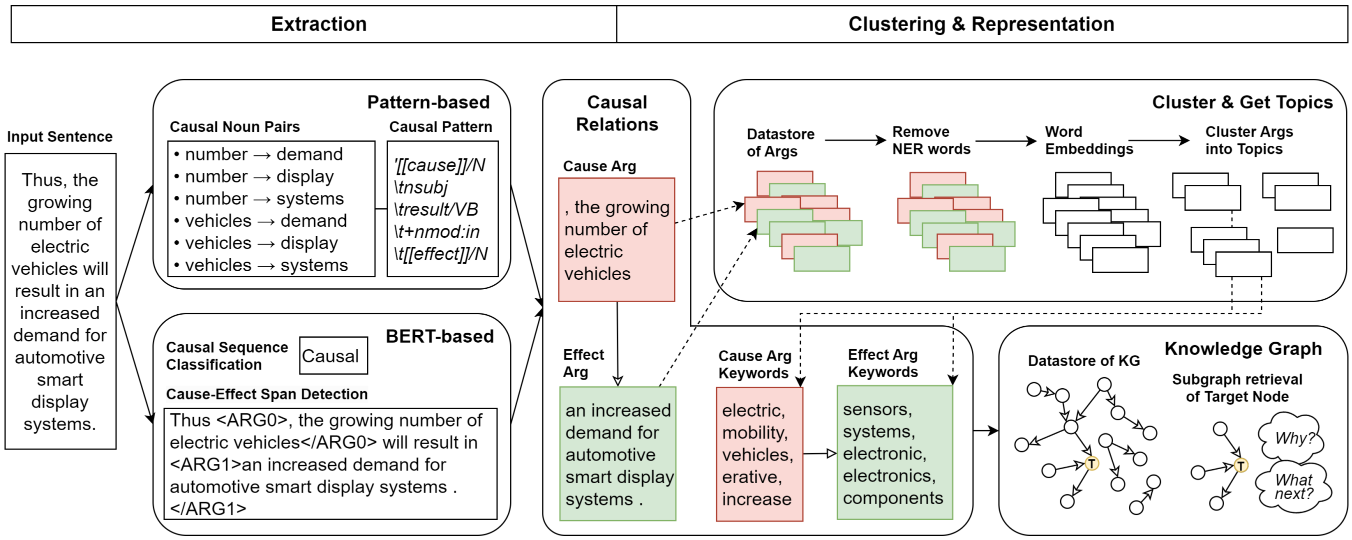

Figure 1: Overview of our methodology.

Compared to pattern-based methods, neural network-based approaches can be trained to recognize more causal constructions, and therefore, in application, have a much higher recall. Yet, many papers working on constructing causal KGs still revert to rudimentary pattern-based solutions (Ittoo and Bouma 2013; Heindorf et al. 2020; Radinsky, Davidovich, and Markovitch 2012; Izumi and Sakaji 2019; Xu and Dang 2022). Recall is important in our context of monitoring news: our users have to be aware of latest causal events, and a pattern-based extraction tool with limited coverage will not be effective in identifying a high proportion of the true causal relations. In our work, we employ both pattern (Heindorf et al. 2020) and neural network-based (Tan, Zuo, and Ng 2023) methodologies based on previous works. We investigate the differences in quality and quantity of the extracted causal relations using these two extraction approaches.

KGs can serve as a taxonomy or knowledge source to guide natural language models to make better predictions (He et al. 2021; Zhang et al. 2021; Cao et al. 2021). Most causal KGs are built off the extracted Cause and Effect arguments by casting them directly as nodes (Heindorf et al. 2020; Izumi and Sakaji 2019; Hassanzadeh 2021). If we followed suit, we will obtain a large and poorly connected graph. In studying causality, it is beneficial to have a highly connected graph because it allows us to detect more causal relations, especially transitive ones. Additionally, generalizing over objects, actions and events allow users to make predictions of upcoming Effects even for unseen events (Radinsky, Davidovich, and Markovitch 2012). Therefore, in our work, we condense our graphs by grouping nodes that refer to the same topic together using previous topic modelling solutions (Sia, Dalmia, and Mielke 2020; Zhang et al. 2022).

Dataset

We worked on 6,384 article summaries, comprising of 62,151 sentences published between 2017 and 2022. We focus on the electronics and supply-chain industry news. The articles were extracted through Google News using a web-scraping tool, `Scrapy`[1] on September to October 2022. We focused on the Japan, China, Europe and Global regions. The article summaries and titles were obtained using `newspaper3K`[2], which returns top 10 sentences of an article, scored and ranked using features such as sentence length, sentence position, title status, and frequency of keywords appearing in the sentence.

Methodology

To briefly introduce, our approach is to extract causal relations, cluster semantically similar arguments, and store causal relations in a KG to be used for various applications.

Extraction of Causal Relations

Pattern-based We replicated CauseNet's (Heindorf et al. 2020) methodology of using linguistic patterns to detect causal relations. The patterns identify the shortest path between a Cause noun and an Effect noun using the dependency graph of a sentence (Culotta and Sorensen 2004; Bunescu and Mooney 2005; Ittoo and Bouma 2013). The enhanced dependency graphs were obtained using the Stanford NLP Parser (Chen and Manning 2014; Schuster and Manning 2016). The original authors extracted 53 linguistic patterns after two bootstrapping rounds on their Wikipedia dataset which we used directly.[3]

To obtain more patterns, we utilized the Wikipedia dataset from (Heindorf et al. 2020)[4], which contains 1,168,155 causal sentences with the Cause and Effect arguments that

[1]https://docs.scrapy.org/en/latest/index.html

[2]https://newspaper.readthedocs.io/en/latest/

[3]We had to alter 28 patterns slightly to fit the our version of Stanford Parser (1.4.1 version) because (Heindorf et al. 2020) used an outdated 0.2.0 version.

[4]https://github.com/causenet-org/CIKM-20

Original	Pattern-based Extraction	Post-processing	Pre-processing for Clustering
... the impact of a fall in output brought on by a global chip shortage.	• shortage → impact • shortage → fall • shortage → output Pattern: '[[cause]]/N -nmod:by brought/VBN +nmod:of [[effect]]'	... the <ARG1>impact of a fall in output</ARG1> brought on by a global chip <ARG0> shortage </ARG0>.'	... the <ARG1>impact of a fall in output</ARG1> brought <ARG0>on by a global chip shortage</ARG0>.'

Table 1: Processing of pattern-based predictions.

were identified by the template-based method. We reverse-engineered the linguistic pattern connecting the Cause and Effect. Out of the 261,643 unique patterns we obtained, we retained the top 50 most common patterns. For the 51st to 500th pattern, we dropped patterns that have no center token (E.g. "[[cause]]/N -nmod:of [[effect]]/N") because such patterns that directly relate the dependency of a potential Cause to a potential Effect will return many spurious, misidentified causal relations. This reverse-engineering method provided us with 477 patterns.

Finally, the original 53 patterns were merged with the additional 477 patterns. Since 43 patterns were repeated across the two lists, the final number of patterns was 487. Most of these patterns contain causal connectives like *'caused'*, *'causing'*, *'resulted in'* and *'leading to'*. Equipped with these linguistic patterns, we extracted causal relations from news as follows:

1. Extract all nouns in a sentence. We use Stanford NLP parser to obtain these part-of-speech (POS) tags.
2. For every combination of noun pairs, identify the shortest dependency path tying the two nouns together. Format the path as a pattern string.
3. Check if the pattern string matches with any of our causal linguistic patterns. If there is a match, the noun pair is identified to be causal.

Post-processing We merged arguments that have the same pattern and either Cause or Effect argument, since they refer to the same relation. To illustrate, in Table 1, the three Effect has the same Cause and the same causal pattern. Therefore, the final causal relation was processed to be *"shortage"* caused *"impact of a fall in output"*. Similarly, in Figure 1, the pattern-based example's six causal relations was simplified into one causal relation: *"number of vehicles"* caused *"demand for automotive smart display systems"*.

Pattern-based arguments tend to be short and lack the context needed for clustering. Therefore, we converted the arguments from the pattern-based approach into words from the original span up to (if occurring before) or until (if occurring after) the signal words. For example, in Table 1, the Cause argument was altered to start right before the signal word *'brought'*. Since the Effect argument already spans up to the signal word, it remains the same.

To conclude this subsection, the pattern-based matching approach allowed us to identify 1,006 sentences and 975 causal relations from 611 unique sentences.

BERT-based UniCausal (Tan, Zuo, and Ng 2023)[5] is a causal text mining repository that consolidated six datasets (AltLex (Hidey and McKeown 2016), BECAUSE 2.0 (Dunietz, Levin, and Carbonell 2017), CausalTimeBank (Mirza et al. 2014; Mirza and Tonelli 2014), EventStoryLine (Caselli and Vossen 2017), Penn Discourse Treebank V3.0 (Webber et al. 2019) and SemEval2010Task8 (Hendrickx et al. 2010)) for three tasks (Causal Sentence Classification, Causal Pair Classification, and Cause-Effect Span Detection). Pre-trained models were created and made available online. These models were trained on each task independently, builds on BERT-based pre-trained encoders (Devlin et al. 2019), and used Cross Entropy Loss. All six datasets were used for training and testing. In our work, we utilized three pre-trained models developed by UniCausal:

1. **Causal Sentence Classification (CSC):** Model identifies if a sentence contains causal relations or not. After passing the sentence through BERT-encoder layers, the embeddings of the `[CLS]` token are processed through a dropout layer, followed by a classification layer to generate predicted logits. The pre-trained model reported 70.10% Binary F1 score.
2. **Causal Pair Classification (CPC):** Model identifies if a pair of arguments (`ARG0`, `ARG1`) that are marked in the sentence are causally related or not, such that `ARG0` causes `ARG1`. It follows the same architecture as CSC. The pre-trained model reported 84.68% Binary F1 score.
3. **Cause-Effect Span Detection (CESD):** Model identifies the consecutive span of words that refer to the Cause and Effect arguments. Framed as a token classification task, after the BERT-encoder layers, the sequence output is fed through a dropout then classification layer to obtain the predicted logits per token. The pre-trained model reported 52.42% Macro F1 score.

To extract causal relations from text, we applied both CSC and CESD predictors to all sentences. For causal sentences identified by CSC, we retained the cause and effect arguments identified by CESD.

Post-processing One limitation of UniCausal's CESD is that it was designed to predict only one causal relation per example. However, in our investigations, many instances had multiple `ARG0` and `ARG1` predictions. Without additional information, the relationship between the multiple causes and effects was unclear. Therefore, we implemented a post-processing procedure involving three steps: (1) Merge sequential arguments, (2) Keep longest argument for examples with three arguments, and (3) Keep multiple causal relations based on CPC predictions. Additionally, we utilized CPC to identify and retain additional causal examples. Details about these procedures can be found in the Appendix.

[5]https://github.com/tanfiona/UniCausal

Altogether, the BERT-based method identified 19,250 sentences with 19,192 causal relations from 15,702 unique sentences.

Argument Clustering

We wish to cluster the arguments that have similar meaning, both in terms of the topic mentioned in the argument (E.g. supply, profits, automotives, etc.) and the impact on it (E.g. positive, negative, etc.). We used the approach by (Sia, Dalmia, and Mielke 2020; Zhang et al. 2022) to generate word embeddings from sequences and cluster the embeddings directly.

Neutralizing named-entities We are not interested to cluster arguments that refer to the same organization, location, or date. Thus, we used the 7-class Stanford Named Entity Recognition (NER) Tagger (Finkel, Grenager, and Manning 2005) [6] to extract named-entities for locations, persons, organizations, times, money, percents, and dates. Subsequently, we remove the words corresponding to any of these entities in the argument spans. For example, the Cause argument *"to jointly produce premium EVs in China"* was converted to *"to jointly produce premium EVs in"*. Note that in the final KG, the original arguments were used.

Word embeddings To generate word embeddings that clusters semantically similar arguments together and semantically different arguments apart, we used the supervised pre-trained language model by SimCSE (Gao, Yao, and Chen 2021) to encode our NER-neutralized arguments into embeddings. SimCSE was trained to identify whether the relationship betweent two sentences suggests entailment, neutral, or contradition. SimCSE was evaluated against standard semantic textual similarity tasks, and achieved an average 81.6% Spearman's correlation, a 2.2% improvement compared to previous best results. Our embeddings had a feature dimension of 786 because the model is built on the BERT model, `bert-base-uncased`.

Clustering and getting keywords per cluster Similar to (Sia, Dalmia, and Mielke 2020), we used K-Means to cluster the 35,230 embeddings into 3,000 topics. We remove relations where the Cause and Effect fall under the same topic so that we do not have nodes that connected to itself. To obtain the top keywords per topic, we used the TFIDF $\times$ IDF method proposed by (Zhang et al. 2022):

$$TFIDF_d = \frac{n_w}{\sum_w n_w} \cdot log(\frac{|D|}{|\{d \in D | w \in d\}|}) \quad (1)$$

$$IDF_k = log(\frac{|K|}{|\{w \in K\}|}) \quad (2)$$

$$TFIDF \times IDF = TFIDF_d \cdot IDF_k \quad (3)$$

where n_w is the frequency of word w in each document d, D refers to the whole corpus, $|K|$ is the number of clusters, and $|\{w \in K\}|$ is the number of clusters where word w appears in. This approach helps to identify the important words to each cluster compared to the rest of the dataset ($TFIDF_d$), while penalizing frequent words that appear in multiple clusters (IDF_k). Empirical findings demonstrate this method significantly outperforms regular TF or TFIDF methods in selecting topic words (Zhang et al. 2022). In the end, we extract 5 keywords per topic, which corresponds to the text displayed in the nodes of a graph. If the cluster contains only one argument, then the argument text is displayed instead.

Knowledge Graph

We define our knowledge graph $G = (V, E)$ as a collection of nodes $V = \{(v_1, v_2, ..., v_n)\}$ and directed edges $E = \{(v_1, v_2), (v_2, v_3), ...\}$. A directed edge (v_i, v_j) represents the presence of causality between the two nodes, where v_i is the Cause and v_j is the Effect. The edges are also weighted by s, which represents the number of sentences in our dataset that has been identified to convey that v_i causes v_j.

Table 2 shows the statistics of our extracted relations and constructed KG. Earlier, out of 62,151 sentences, we identified 15,902 unique sentences containing 20,086 causal relations. Before clustering, a KG built directly on these extracted relations would have 35,230 unique nodes and 20,086 unique edges, with an average support per edge of 1.008. By performing argument clustering, we created a highly connected KG, with 3,000 unique nodes, 17,801 unique edges, and an average support per edge of 1.122. Table 3 displays some graph statistics before and after clustering. Again, we observe that the KG after clustering is denser and more connected. In fact, instead of 15,686 subgraphs, our KG is now represented by 1 connected graph. Visualizations of the KGs in the later sections are performed using Cytoscape[7], an open-source software for visualizing and interacting with graphs.

Experimental Results

Extraction of Causal Relations

Quantitative evaluation We asked users to randomly select 15 articles from the Google News dataset and annotate the Causes and Effects per sentence. 49 causal relations from 43 sentences were identified. 6 sentences had more than one causal relation. A correct case is one where the model and human annotations have ≥ 1 word(s) overlapping for both Cause and Effect spans. We could then calculate the number of True Positives (TP), False Positives (FP) and False Negatives (FN) by treating the human annotations as the gold standard. True Negative (TN) is 0 in all cases because we do not evaluate on non-causal sentences and relations. Appendix Table **??** provides some examples comparing model predictions to human annotations. Subsequently, we calculate Precision (P), Recall (R), and F1 scores using the following formulas:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F1 = \frac{2 \times P \times R}{P + R} \quad (4)$$

[6] https://nlp.stanford.edu/software/CRF-NER.shtml

[7] https://cytoscape.org/

Extraction					Before Clustering			After Clustering		
Method	$\|Sents\|$	$\|Sents\|$	$\|Rels\|$	Avg Rel Support	$\|V\|$	$\|E\|$	Avg E Support	$\|V\|$	$\|E\|$	Avg E Support
Pattern-based	1,006	611	975	1.032	1,476	975	1.032	774	845	1.340
BERT-based	19,250	15,702	19,192	1.003	33,940	19,192	1.003	2,990	17,075	1.120
Total	20,255	15,902	20,086	1.008	35,230	20,086	1.008	3,000	17,801	1.122

Table 2: Summary statistics of extracted causal relations per step.

	Before	After
No. of Unique Nodes, $\|V\|$	35,230	3,000
No. of Unique Edges, $\|E\|$	20,086	17,801
Total Weight, $\sum s$	20,254	19,965
No. of Subgraphs	15,686	1
Avg Clustering Coefficient	$9.81e^{-06}$	$1.75e^{-02}$
Avg Degree Centrality	$3.24e^{-05}$	$3.96e^{-03}$
Avg Eigenvector Centrality	$6.64e^{-05}$	$1.32e^{-02}$
Transitivity	$4.17e^{-04}$	$8.81e^{-03}$

Table 3: Graph statistics before and after clustering.

Extraction Method	P	R	F1
Pattern-based	**100.00**	4.08	7.84
BERT-based	76.09	71.43	73.68
Both	75.00	**73.47**	**74.23**

Table 4: Performance metrics for extraction on human-annotated test set. Scores are reported in percentages (%). Top score per column is bolded.

Our model identified 48 causal relations from 43 sentences. 4 sentences had more than one causal relation. The performance metrics are reported in Table 4. By using our proposed method of combining both pattern-based and BERT-based approaches in extraction, the F1 score is the highest at 74.23%.

Although the pattern-based has very good precision (100%), it could only identify 2/49 causal relations, resulting in an extremely low recall score of 4%. In the application of monitoring news and trends, such a low recall is unacceptable as it would miss out on many key happenings. The BERT-based approach extracted much more causal relations than the pattern-based method, scoring a higher recall of 71.43%. This is because UniCausal is trained on a large dataset, and its architecture also allows models to learn from varied linguistic structures, including implicit causal relations. Consistent with our findings, we observe in Table 2 that the BERT-based approach extracted 19x more causal relations than the pattern-based approach for our whole dataset.

The model proposed 12 causal relations that were not annotated by the humans. Upon checking, 5/11 were correct in that the Cause and Effect spans suggested are causal. However, they are duplicates arising from the post-processing done for BERT-based extraction that accepts any pair of arguments that CPC detects as causal. Therefore, when comparing against the human annotated test set, these duplicates were treated as spurious relations. If we consider these five examples as Correct, then the overall precision would increase to 85.42%.

Argument Clustering

Quantitative evaluation From the 36 causal relations where the model and users were in agreement with, we asked users to group the arguments with similar meaning and give a topic label to each group. The users clustered 72 arguments into 50 topics. Some example topics are: *'increased competition'*, *'taxation'*, *'cost reduction'*, *'business hurdles'* and *'raw material shortage'*. To obtain the model's predictions, we filtered out the nodes of the 36 causal relations from the whole KG (described in the "Knowledge Graph" Section). The 72 arguments were clustered into 70 topics. We compare the model's and user's clustering using Normalized Mutual Information (NMI), an entropy-based evaluation metric. Because most clusters only contain one span from both the model's and user's clustering, NMI is high at 93.62%. However, given the small sample size, this score can be misleading. More annotated data is needed to evaluate clustering performance.

Qualitative evaluation Due to limited space, we summarize our qualitative experiments and findings in this subsection. Details are available in the Appendix.

In Table 3, we show that clustering helps to increase the average edge weight and node centrality. Our argument clustering solution helps to create a highly connected causal KG, which is more insightful to infer causal relationships from. For example, before clustering, we could only infer that *'pandemic'* causes *'disruptions'*. After clustering, our subgraph detected that *'pandemic'* causes supply chain disruptions, chip and semiconductor shortages, sales decreases, and general interferences and disruptions.

We also found that argument clusters are much more defined on a 2D plot after we remove the named-entities from arguments. Named-entity removal allows the clustering process to focus on more meaningful words referring to the event, sentiment, or topic instead.

Applications in the Industry

Use Cases

Summarization Our causal KG is useful for summarizing reported causal relations in news. In the earlier section about qualitative evaluation, we constructed a *'pandemic'* subgraph and demonstrated how we can swiftly learn about the reported effects of pandemic.

Answering causal questions and predicting future events Our KG is also useful for answering causal questions. In Figure 1, users that learn that that Event A (*"the growing number of electric vehicles"*) causes Event B (*"an increased demand for automotive smart display systems"*) might ask: What might happen next as a result of Event B? By setting the target node to be Event B, we identify that the next two likely events (based on edge support) are that display systems will *"become an integral part of the automotive supply chain"* (*'automotive_industry_shaft_xa_siness'*) and that this *"trend is expected to continue during the forecast period"* (*'forecast_period_during_anticipated_analysts'*). Other subsequent Effects are: *"so does demand for wiring harnesses and related electronic sub-assemblies"*, *"forcing European car makers to rely on Asian suppliers"*, *"the task of designing today's cars much more difficult"*, and many more meaningful predictions. We can also ask other causal questions like "Are there other causes of Event B?" and "What caused Event A in the first place?"

Inferring transitive causal relations In transitive relations, if Event A causes Event B, and Event B causes Event C, then Event A can also be said to cause Event C. For causal relations, the transitive property fails if the relations are too specific[8]. Referring to the example from the earlier paragraph, we observe that transitivity does hold: *"the growing number of electric vehicles"* (Event A) does help make automotive smart display systems *"become an integral part of the automotive supply chain"* (Event C) through *"an increased demand for automotive smart display systems"* (Event B). This observation has strong implications on how insightful our KG can be for inferring causal relationships that were not otherwise stated directly in the news. Further analysis will be needed to identify the cases where transitivity fails, and how we should handle them.

Trend monitoring We demonstrate the potential of our KG to reflect trends over time. To conduct the experiment, we split our dataset into articles that are published across three time baskets: Before 2020, 2020 to 2021 (inclusive), and after 2021. Since our dataset is very imbalanced across time, we down-sampled the two larger baskets such that all baskets have the same sample size. Similar to analyses before, we created subgraphs by filtering out target nodes and nodes that are one step away from target node(s). A node is a target if the search term(s) can be found in the node description in any order. In Figure 2, we studied three search terms across the three time baskets. For each subgraph, the edges highlighted in red falls within the time basket of interest. Our findings show that the frequency of causal relationships about the topic *"chip shortage"* was rare before 2020, extremely heated during the pandemic period of 2020 to 2021, and lower from 2021 onwards. This is validated by experts' understanding that the COVID-19 pandemic kick-started the chip shortage, amongst many other reasons. As a sanity check, we found that no causal relationships were mentioned about *"pandemic"* before 2020. This makes sense because the awareness of COVID-19 only started taking off in the first quarter of 2020. As a control, we also checked that interest in topics about *"robotics"* stayed roughly constant throughout the three time baskets. To conclude, our KG can be helpful for monitoring heated causal topics and news trends across time.

[8]Example of specific causal relations violating transitivity: *"Sugar makes John happy. Sugar causes diabetes. Diabetes makes John sad. Does sugar make John happy or sad?"*

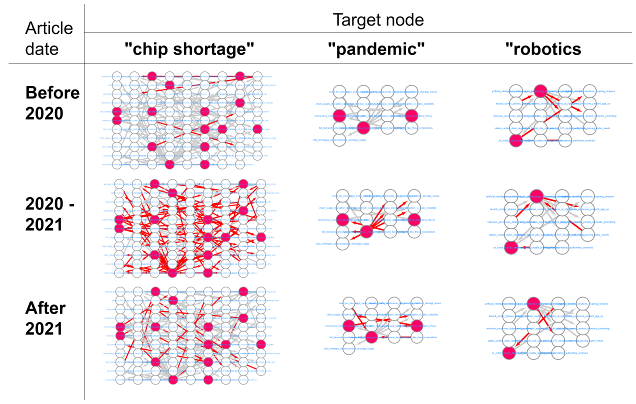

Figure 2: Subgraph(s) filtered based on nodes that are one step away from target node(s) highlighted in pink. In this example, a node is referred to as a target if it contains the search term in any order indicated on top (E.g. *'chip shortage'*). Edges from articles falling under the time period corresponding to the right axis are highlighted in red.

User Feedback

The final KG was presented users to gather feedback. Response was positive, and many users could find a use for this KG in their daily work. We intend to deploy the our system to generate regular snapshots of the news that will serve as market reports for the Sales Division. Users believe that harnessed with knowledge about recent causal events, they can improve their market understanding and perform better at prediction-related tasks, like sales and demand forecasting. In the future, users would also like to see the KG to be improved by adding temporal and sentiment elements.

Conclusion & Future Work

We focused on the application of extracting causal relations in industry news to construct a causal KG. Our approach was (1) recall-focused, by employing BERT-based on top of pattern-based extraction methods. Our approach was also (2) precision-focused, and for our test set, achieved 75% score. Finally, our final KG was designed to be (3) interpretable, with many use cases and was deemed useful by our users in the industry. Our work can be replicated onto many other domains. In the future, we intend to annotate a larger test set for more concrete evaluation. Additionally, we plan to deploy the extraction and graphing system to generate monthly snapshots of the market. We will monitor the users' interactions with the system and identify areas for improvement. Lastly, we hope to include temporal and sentiment elements in our extraction to enrich the final KG.

References

Aziz, A.; Hossain, M. A.; and Chy, A. N. 2022. CSECU-DSG @ Causal News Corpus 2022: Fusion of RoBERTa Transformers Variants for Causal Event Classification. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, 138–142. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics.

Bunescu, R.; and Mooney, R. 2005. A Shortest Path Dependency Kernel for Relation Extraction. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 724–731. Vancouver, British Columbia, Canada: Association for Computational Linguistics.

Cao, P.; Zuo, X.; Chen, Y.; Liu, K.; Zhao, J.; Chen, Y.; and Peng, W. 2021. Knowledge-Enriched Event Causality Identification via Latent Structure Induction Networks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 4862–4872. Online: Association for Computational Linguistics.

Caselli, T.; and Vossen, P. 2017. The Event StoryLine Corpus: A New Benchmark for Causal and Temporal Relation Extraction. In *Proceedings of the Events and Stories in the News Workshop*, 77–86. Vancouver, Canada: Association for Computational Linguistics.

Chen, D.; and Manning, C. 2014. A Fast and Accurate Dependency Parser using Neural Networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 740–750. Doha, Qatar: Association for Computational Linguistics.

Chen, M.; Cao, Y.; Deng, K.; Li, M.; Wang, K.; Shao, J.; and Zhang, Y. 2022a. ERGO: Event Relational Graph Transformer for Document-level Event Causality Identification. In *Proceedings of the 29th International Conference on Computational Linguistics*, 2118–2128. Gyeongju, Republic of Korea: International Committee on Computational Linguistics.

Chen, X.; Zhang, G.; Nik, A.; Li, M.; and Fu, J. 2022b. 1Cademy @ Causal News Corpus 2022: Enhance Causal Span Detection via Beam-Search-based Position Selector. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, 100–105. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics.

Culotta, A.; and Sorensen, J. 2004. Dependency Tree Kernels for Relation Extraction. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, 423–429. Barcelona, Spain.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. 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)*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.

Dunietz, J.; Levin, L.; and Carbonell, J. 2017. The BECauSE Corpus 2.0: Annotating Causality and Overlapping Relations. In *Proceedings of the 11th Linguistic Annotation Workshop*, 95–104. Valencia, Spain: Association for Computational Linguistics.

Finkel, J. R.; Grenager, T.; and Manning, C. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, 363–370. Ann Arbor, Michigan: Association for Computational Linguistics.

Gao, T.; Yao, X.; and Chen, D. 2021. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 6894–6910. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Hassanzadeh, O. 2021. Building a Knowledge Graph of Events and Consequences Using Wikidata. In Kaffee, L.; Razniewski, S.; and Hogan, A., eds., *Proceedings of the 2nd Wikidata Workshop (Wikidata 2021) co-located with the 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24, 2021*, volume 2982 of *CEUR Workshop Proceedings*. CEUR-WS.org.

He, L.; Zheng, S.; Yang, T.; and Zhang, F. 2021. KLMo: Knowledge Graph Enhanced Pretrained Language Model with Fine-Grained Relationships. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, 4536–4542. Punta Cana, Dominican Republic: Association for Computational Linguistics.

Heindorf, S.; Scholten, Y.; Wachsmuth, H.; Ngomo, A. N.; and Potthast, M. 2020. CauseNet: Towards a Causality Graph Extracted from the Web. In d'Aquin, M.; Dietze, S.; Hauff, C.; Curry, E.; and Cudré-Mauroux, P., eds., *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, 3023–3030. ACM.

Hendrickx, I.; Kim, S. N.; Kozareva, Z.; Nakov, P.; Ó Séaghdha, D.; Padó, S.; Pennacchiotti, M.; Romano, L.; and Szpakowicz, S. 2010. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, 33–38. Uppsala, Sweden: Association for Computational Linguistics.

Hidey, C.; and McKeown, K. 2016. Identifying Causal Relations Using Parallel Wikipedia Articles. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1424–1433. Berlin, Germany: Association for Computational Linguistics.

Ittoo, A.; and Bouma, G. 2013. Minimally-supervised learning of domain-specific causal relations using an open-domain corpus as knowledge base. *Data Knowl. Eng.*, 88: 142–163.

Izumi, K.; and Sakaji, H. 2019. Economic Causal-Chain Search using Text Mining Technology. In *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, 61–65. Macao, China.

Mirza, P.; Sprugnoli, R.; Tonelli, S.; and Speranza, M. 2014. Annotating Causality in the TempEval-3 Corpus. In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, 10–19. Gothenburg, Sweden: Association for Computational Linguistics.

Mirza, P.; and Tonelli, S. 2014. An Analysis of Causality between Events and its Relation to Temporal Information. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2097–2106. Dublin, Ireland: Dublin City University and Association for Computational Linguistics.

Nik, A.; Zhang, G.; Chen, X.; Li, M.; and Fu, J. 2022. 1Cademy @ Causal News Corpus 2022: Leveraging Self-Training in Causality Classification of Socio-Political Event Data. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, 91–99. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics.

Radinsky, K.; Davidovich, S.; and Markovitch, S. 2012. Learning causality for news events prediction. In Mille, A.; Gandon, F.; Misselis, J.; Rabinovich, M.; and Staab, S., eds., *Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, April 16-20, 2012*, 909–918. ACM.

Schuster, S.; and Manning, C. D. 2016. Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2371–2378. Portorož, Slovenia: European Language Resources Association (ELRA).

Sia, S.; Dalmia, A.; and Mielke, S. J. 2020. Tired of Topic Models? Clusters of Pretrained Word Embeddings Make for Fast and Good Topics too! In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1728–1736. Online: Association for Computational Linguistics.

Tan, F. A.; Zuo, X.; and Ng, S.-K. 2023. UniCausal: Unified Benchmark and Repository for Causal Text Mining. In *Big Data Analytics and Knowledge Discovery*. Springer International Publishing.

Webber, B.; Prasad, R.; Lee, A.; and Joshi, A. 2019. The penn discourse treebank 3.0 annotation manual. *Philadelphia, University of Pennsylvania*.

Xu, Z.; and Dang, Y. 2022. Data-driven causal knowledge graph construction for root cause analysis in quality problem solving. *International Journal of Production Research*, 1–19.

Zhang, Z.; Fang, M.; Chen, L.; and Namazi Rad, M. R. 2022. Is Neural Topic Modelling Better than Clustering? An Empirical Study on Clustering with Contextual Embeddings for Topics. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3886–3893. Seattle, United States: Association for Computational Linguistics.

Zhang, Z.; Wang, H.; Zhao, H.; Tong, H.; and Ji, H. 2021. EventKE: Event-Enhanced Knowledge Graph Embedding. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, 1389–1400. Punta Cana, Dominican Republic: Association for Computational Linguistics.

Zuo, X.; Cao, P.; Chen, Y.; Liu, K.; Zhao, J.; Peng, W.; and Chen, Y. 2021a. Improving Event Causality Identification via Self-Supervised Representation Learning on External Causal Statement. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2162–2172. Online: Association for Computational Linguistics.

Zuo, X.; Cao, P.; Chen, Y.; Liu, K.; Zhao, J.; Peng, W.; and Chen, Y. 2021b. LearnDA: Learnable Knowledge-Guided Data Augmentation for Event Causality Identification. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 3558–3571. Online: Association for Computational Linguistics.

Zuo, X.; Chen, Y.; Liu, K.; and Zhao, J. 2020. KnowDis: Knowledge Enhanced Data Augmentation for Event Causality Detection via Distant Supervision. In *Proceedings of the 28th International Conference on Computational Linguistics*, 1544–1550. Barcelona, Spain (Online): International Committee on Computational Linguistics.