Automatic Learning and Reasoning of Causal Knowledge in the Financial Domain

Abstract

Causal reasoning is one of the keys to reach artificial intelligence. To achieve it, two focal challenges are how to design an effective causal knowledge representation scheme to reason and how to automatically obtain this kind of causal knowledge on a large scale. We try to counter these challenges in this paper. First, we design a novel causal knowledge representation scheme based on the logic rule, which takes structured events as the basic key components and can reason with uncertainty. Second, we propose a rule learning framework that automatically acquires a large number of rules from unstructured text. Last, the experiments demonstrate the power of our proposed causal knowledge representation scheme and the rules learned are satisfactory in both qualitative and quantitative aspects.

1 Introduction

Causal reasoning, the core challenge in artificial intelligence, which aims to understand the causal dependency between events, is receiving more and more attention [?]. In the reasoning process, causal knowledge plays a critical role in people's daily behavior and decision-making [?]. It is of great interest in many domains, including finance, where understanding causal relationships can provide significant opportunities for economic benefits. For example, consider the following,

1. If a large disaster happens in a country and this country is rich in certain metal, the price of this metal will rise.
2. If the price of some kind of metal rises, the price of the products based on this metal will also go up.

Above causal knowledge expressed in the form of natural language is easy for us to understand and has great practical value in real life through simple reasoning. For example, if there's an earthquake in Chile and we know above rule 1, it is easy to infer the price of copper will rise. Further, we can infer the price of the household appliances, such as air conditioners and refrigerators, will also rise via above rule 2.

However, it is a daunting task for humans, especially traders, to learn many of these rules and use them for real-time reasoning in the real world. We hope machines can learn these rules automatically and reason quickly with them to help us get rid of the heavy burden of rule learning and real-time massive information processing. In order to achieve this goal, we face two challenges: how to represent causal knowledge in a machine-actionable way and how to automatically acquire a large amount of this type of causal knowledge. We separately explain these two aspects.

1.1 Causal Knowledge Representation and Reasoning

To get some inspirations on how to represent causal knowledge, we first review some previous causal knowledge representation schemes. Now, there are two general directions about causal knowledge representation. One direction is ***Neural or numeric*** form. Such neural form scheme can represent both causal knowledge and non-causal knowledge by a unified graph-embedding method [?; ?]. However, it not only has the problems of interpretability and reusability but also has the problem of weak reasoning ability. The other direction is the ***symbolic*** form, which also includes two directions. One is the specific description of causal knowledge. Both the cause and the effect of the causal knowledge are specific events or actions, expressed by terms or short text, such as CausalNet[?], ConceptNet[?]. However, this kind of knowledge is usually less expressive and less informative. The other is the general description of causal knowledge. Both the cause and the effect of the causal knowledge are abstract events or actions, expressed by structures, such as frame [?] or a pair of abstract words [?]. These structures are usually too abstract or vague to be understood. Additionally, existing symbolic representation schemes are unfriendly for machines to reason.

Previous causal knowledge representation schemes encounter problems, such as uninterpretability, insufficient expression ability, unfriendly reasoning. Therefore, we propose a novel and powerful representation scheme with logical form, which takes structured events as the basic key components and can reason with uncertainty. Here, we use the following logic rule (1), equivalent to above natural language rule 1, as an example to illustrate how this idea evolved. (Z,'价格/price','上涨/rise','',''):-('',X,'遭受/suffer',Y,'袭击/attack'),isA(X,'国家/country'),isA(Y,'自然灾害/ disaster'),isA(Z,'金属/metal'),atLocation(Z,X) conf:0.842 (1)

First, we try to use structures to represent the events in the cause part and effect part of the causality well. A simple and effective symbolic event representation is a structured form *(Subject, Predicate, Object)* as used in [?] and [?], also known as SPO. We extend the SPO form to *(Modifier of Subj, Subj, Predicate, Modifier of Obj, Obj)* to capture richer event information. We call this event pair a rule instance, see (2) (following the convention of Prolog, we put the cause part in the head position).

(‘铜/copper’,‘价格/price’,‘上涨/rise’,‘’,‘’):-(‘’,‘智利/Chile’,‘遭受/suffer’,‘地震/earthquake’,‘袭击/attack’) (2)

Second, using such a specific rule instance (2) to represent causal knowledge is less informative and lacks the ability to infer the unseen. Instead, we hope it's a general rule (3).

(Z,‘价格/price’,‘上涨/rise’,‘’,‘’):-(‘’,X,‘遭受/suffer’,Y, ‘袭击/attack’),isA(X,‘国家/country’),isA(Y,‘自然灾害/disaster’),isA(Z,‘金属/metal’) (3)

Third, reasoning using rule (3) may result in some errors. For example, we can get one inferred result: (‘铁/steel’,‘价格/price’,‘上涨/rise’,‘’,‘’):-(‘’,‘智利/Chile’, ‘遭受/suffer’, ‘地震/earthquake’,‘袭击/attack’). This is unreasonable because Chile is a large copper producer rather than a large steel producer. Therefore, we add some constraint relations, such as atLocation(Z,X), to exclude unreasonable inference.

Last, after previous step, we assign a confidence value to enable it to reason with uncertainty. Finally, we get rule (1).

To sum up, this kind of rule is very informative and friendly for machines to reason. By the way, our causal knowledge representation scheme can be easily extended to the conjunction of multi-cause events by adding multiple cause events to the rule header.

1.2 Causal Knowledge Acquisition

After developing a machine-actionable causal knowledge representation scheme with the logic rule, we need to tackle the second challenge about how to automatically obtain a large number of rules. WWW (World Wide Web) is a very large treasure trove of knowledge, but most of the content is unstructured and noisy text, which challenges us to learn rules. Rule learning has been studied extensively in Inductive Logic Programming (ILP) [?; ?]. However, the rule instances extracted from Web text are noisy and incomplete. That negative examples are mostly absent cannot make the closed-world assumption typically made by ILP systems. This paper presents a new ILP system, which adopts a bottom-up approach with two general stages:

1) From unstructured text to structured rule instances Online text is massive but noisy. We first derive the sentences with causal relation through the designed patterns. Then, we use event extraction technology to extract the structured rule instances, such as above (2).

2) From specific rule instances to general rules With a large number of rule instances, we generalize them into rules, such as above (1), and balance generality and specificity:

Generality We try to represent given rule instances semantically with as few general rules as possible under the help of Probase[?]. For example, induce two rule instances (‘corn/soybean’,‘price’,‘fall’, ‘’,‘’):-(‘corn/soybean’, ‘yield’,‘rise’,‘’,‘’) into the rule (X,‘price’,‘fall’,‘’,‘’) :- (X, ‘yield’,‘rise’,‘’,‘’),isA(X,‘food’).

Specificity Overgeneral rules may lead to errors. For example, inducing two rule instances (‘corn/soybean’,‘price’, ‘fall’,‘’,‘’):-(‘corn/soybean’, ‘yield’,‘rise’,‘’,‘’) into the rule (X,‘price’,‘fall’,‘’,‘’):-(X,‘yield’,‘rise’,‘’,‘’),isA(X,‘thing’) is unreasonable. Because when the machine instantiates ‘thing’ into ‘air’, the inferred events “the yield or price of air rises” do not make sense.

Generalizing rule instances to rules can be seen as a semantic compression procedure, which accords with the MDL (Minimum Description Length) principle. Hypotheses (H) accords with rules and data (D) accords with rule instances. Thus, we propose a rule induction algorithm, which takes MDL as the evaluation criterion and regards this rule induction procedure as an optimization problem, see detail in 2. This procedure is similar to [?] and [?]. Finally, each learned rule will be automatically assigned a confidence value by considering some heuristics.

Contributions. In this paper, we present a full stack solution for causal knowledge representation, acquisition, and reasoning. More precisely, our contributions are as follows: First, we design a novel and powerful causal knowledge representation scheme based on the logic rule with the ability of uncertainty reasoning. Second, we propose a rule learning framework for obtaining rules from large unstructured text and experiments show that the rules leaned are reasonable and effective. Specifically, we learned 50000 rules and human's evaluation shows the top 10000 rules of the highest confidence reach 32.5%(good), 39%(fair), 28.5%(bad). Last, we release the rules and the translated Chinese Probase and ConceptNet and provide an interactive demo to show the reasoning process and the application of futures price triggering.

2 Approach

In this section, we introduce three parts of our proposed rule learning framework: rule instances extraction, rule acquisition, and causal reasoning, as shown in Figure 1.

2.1 Rule Instance Extraction

This part is how to extract rule instances from the text.

Pattern Matching Causality, expressed by natural language texts, can be identified by linguistic patterns known as causal cues [?]. We design a set of causal patterns to extract cause span and effect span from the sentences. Causal patterns can be divided into 3 groups based on the pattern structure, shown in Table 2. The priority of each group represents the order of matching when multiple causal patterns can match a sentence. The stricter the regular expression is, the earlier it matches, and the higher the priority is. In addition, the experiment shows that such kind of patterns can not discover the conjunction of multi-cause events because of people's bad writing habits in the news.

Rule Instance Extraction First, we use these causal patterns to extract cause span and effect span. Then, we further extract the cause events and effect events from these two spans via existing dependency parser [?]. The events are mediated by predicates, so we regard each verb in the span as a predicate. Then, find out the subject and object corresponding to each predicate, and get their modifiers. Meanwhile,

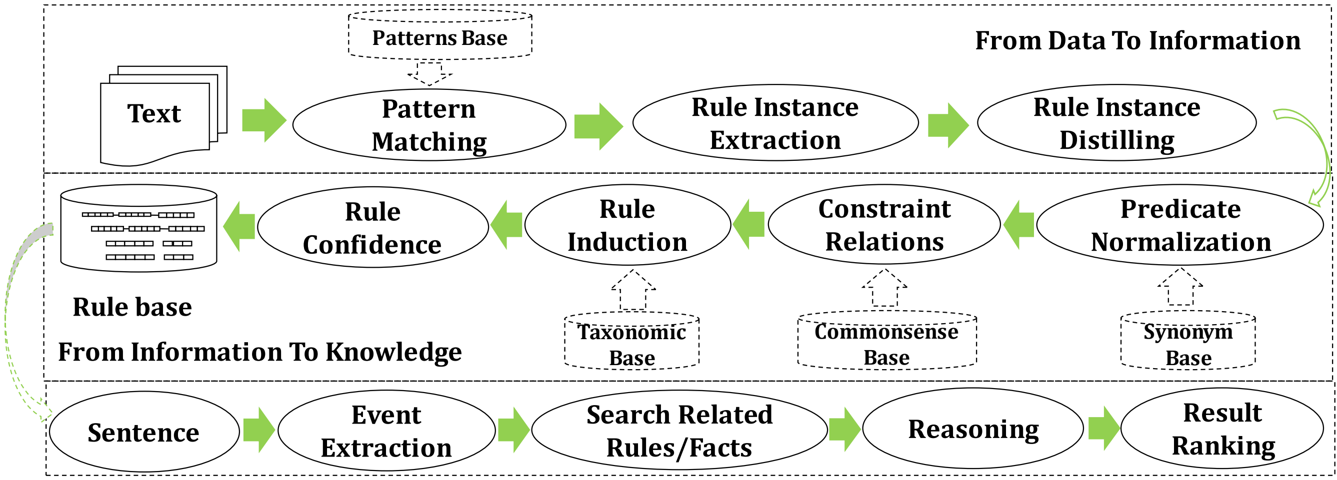

Figure 1: Overview of the framework. The three parts from top to bottom are rule instance extraction, rule acquisition, and reasoning.

we also consider the negative dependency of the verb and mark it behind the predicate, which is a practical implement trick. For example, 'occur_1' indicates 'occur', while 'occur_0' means 'not occur'. Since it may be many predicates in each span, which means many events exist in cause span or effect span, we do a Cartesian product of these cause events and effect events as the extracted rule instances. As the comment of philosopher David Hume about causality that the frequency of causal events will be higher than these are non-causal, filtering will alleviate the noisy rule instances.

Rule Instance Distilling Some heuristics can discard bad rule instances led by event extraction. For example, a) pronouns appearing in events are meaningless. b) some verbs should not be predicates, such as 'say', 'state'. c) complete semantic. An event should have at least one subject or object.

2.2 Rule Acquisition

The part is how to exploit external knowledge bases to generalize rule instances into rules.

Predicate Normalization Predicates in different words may express the same meaning, such as 'raise', 'rise', 'increase'. We use Ciling[1], the largest word-level Chinese synonym resource, to normalize the predicates. We ignore verb disambiguation to reduce the complexity of the framework.

Constraint Relations As analyzed in section 1, we use ConceptNet to add constraint relations, as shown in step d of figure 2. We choose ConceptNet because the objects we care about in finance are always common. Multiple edges between two nodes in ConceptNet are preprocessed to reduce the complexity of our framework. For example, if two relations 'madeof' and 'relateto' exist in two nodes, we will remove 'relateto' based on the semantic richness of the relations. By the way, these constraint relations are not discovered by accident, because if two events are causal, they must be related via some relations.

Rule Induction Generalizing rule instances to rules balances generality and specificity, which is consistent with MDL principle trying to find the best hypothesis H(rules) that can semantically describe or compress data D(rule instances).

a) Generalization We generalize the event roles(subject, object or modifiers) in rule instances using Probase. A pre-built lexicon determines which one can be generalized. For example, it includes concrete things, such as corn, oil, copper, etc., and does not include abstract things, such as price, yield, sale, etc. The more general concepts in Probase a rule has, the more general it is. Formally, let RI and R be the entire rule instances and the rules, ri and r are one of them, i and c are an instance and a concept, respectively.

$$L(D|H) = L(RI|R) = \sum_{r \in R, ri \in RI} p(ri) * (-\log(p(ri|r)))$$

$$p(ri|r) = \prod_{c \ in \ r, i \ in \ ri} p(i|c), \quad p(i|c) = f(c,i)/f(c)$$

where the frequency of c and i pair is $f(c,i)$, obtained directly from Probase, and $p(ri)$ is the frequency of rule instance ri.

b) Specialization As analyzed in 1.2, we try to control the generality of the rules using the entropy of rules. Formally,

$$L(H) = L(R) = \sum_{r \in R} p(r) * (-\log(p(r)))$$

$$p(r) = \sum_{ri \in RI} p(ri)p(r|ri) = \sum_{ri \ s.t. \ ri \ generalizes \ to \ r} p(ri)$$

Balancing generalization and specialization, we will get:

$$R = \arg\min_{R}(\alpha L(R) + (1-\alpha) L(RI|R))$$

Note that the parameter α controls the relative importance of generalization and specialization. Last, we adopt a simulated annealing (SA) algorithm 1 to search the optimal rules (R).

Rule Confidence Knowledge discovered from noisy data is uncertain. Therefore, we combine the following features of the learned rules to give each rule a confidence value. For a given rule, we consider: the frequency of the predicate pair in cause event and effect event, the number of rule instances

[1] http://www.bigcilin.com/

Algorithm 1 Rule Induction

Input: Rule Instances(RI), Translated Probase
Output: Rules(R)

Initialize a rule(r) for each rule instance(ri).
while There is no change of L in the last β iterations **do**
    Generalize a randomly picked rule instance (ri) to a new rule r^{t+1}
    Calculate L^{t+1} $(L = \alpha L(H) + (1-\alpha)L(D|H))$
    Accept this rules(R) with the probability:

$$p = \begin{cases} 1 & L^{(t+1)} < L^{(t)} \\ e^{(L^{(t)} - L^{(t+1)})/t} & L^{(t+1)} >= L^{(t)} \end{cases}$$

end while
Return Rules(R)

that generalizes to this rule, the pattern priority of the rule instances extracted from the sentences, and the frequency of event roles in this rule. Then, get the numeric values of these features of all rules and normalize them to a range of 0 to 1. Finally, the weighted sum is calculated as rule confidence, where the corresponding weight is (0.5, 0.2, -0.1, 0.5).

a) Raw Sentence:
(Chinese) 上个月智利曾遭受8.8级大地震的袭击，导致铜价格上涨近6%。
(English) Copper prices rose nearly 6% last month since Chile was hit by a massive 8.8-magnitude earthquake.
b) Rule Instance Extraction:
('铜/copper', '价格/price', '上涨/rise', '', ''):-('', '智利/Chile', '遭受/suffer', '地震/earthquake','袭击/attack')
c) Predicate Normalization:
('铜/copper', '价格/price', '上涨/rise', '', ''):-('', '智利/Chile', '遭受/suffer', '地震/earthquake','袭击/attack')
d) Constraint Relations:
('铜/copper', '价格/price', '上涨/rise', '', ''):-('', '智利/Chile', '遭受/suffer', '地震/earthquake','袭击/attack'), atLocation(铜/copper,智利/Chile)
e) Rule Induction:
(Z, '价格/price', '上涨/rise', '', ''):-('', X, '遭受/suffer', Y,'袭击/attack'), isA(X,'国家/nation'), isA(Y, '自然灾害/disaster'), isA(Z, '金属/metal'), atLocation(Z,X)
f) Rule Confidence:
(Z, '价格/price', '上涨/rise', '', ''):-('', X, '遭受/suffer', Y,'袭击/attack'), isA(X, '国家/country'), isA(Y, '自然灾害/disaster'), isA(Z, '金属/metal'), atLocation(Z,X) conf:0.842

Figure 2: An Example Showing Various Stages of the Framework

2.3 Causal Reasoning

The specific reasoning process is as follows. Extract the query event from a given query sentence, search related rules in the learned rules with the depth-first search algorithm, look up related facts in Probase and ConceptNet to reduce the size of the facts and rules to speed up reasoning, and convert the found rules and facts into Prolog code together. To implement uncertain reasoning, we put rule confidence into the generated Prolog code and use multiplication to simulate the decline of the confidence in reasoning. Uncertain reasoning should reduce confidence as reasoning deepens. We set the threshold γ to cut off more in-depth reasoning. Small modification of Prolog code generalization is enough to reason with uncertainty, so we choose the mature and reliable Prolog, instead of ProbLog [**?**] and PSL [**?**]. After generating the Prolog code, we query the query event in the Prolog code format and finally return the top K results sorted by the confidence.

3 Experiment Evaluation

In this section, we first give some statistics of our corpus and evaluate the quality and quantity of the learned rules. Then, we compare with other causal knowledge bases. Next, we analyze and discuss some main sub-modules in the rule learning framework. Finally, a practical application of futures price prediction and demonstration are introduced. Our experiments are implemented in Python and SWI-Prolog[2].

3.1 Dataset

We crawled the text dataset from Chinese financial news website [3]. The news data containing 4,991,000 articles, from 2000/7/20 to 2017/12/31, is used to rule learning. The number of sentences with causal cue words is 7,147,141, accounting for 9.46% of the total number of de-duplicated sentences (75,572,053. The repetition rate of sentences is about 32%). It shows that about **14.2%** (9.64%/(1-32%)) sentences explicitly express causality in online financial news sentences. The news data containing 270,562 articles, from 2018/1/1 to 2018/11/2, is used to evaluate our framework. We set α to 0.5 to achieve an equal balance between generalization and specialization in rule induction. We set γ to 0.3 to control the Prolog engine to reason around two steps, since more than two steps lead to obviously unreasonable results.

3.2 Rule Evaluation

We evaluate these rules both quantitatively and qualitatively.

Quantitative Evaluation The number of the final rules we learned is **50000**. We divide the rule quality into three levels: good, fair and bad. According to the ranking of rule confidence, we randomly select 200 rules from the top 10000 rules and manually divide them into three levels. The 'good', 'fair', and 'bad' levels of them account for **32.5%**, **39.0%** and **28.5%**, respectively.

Qualitative Evaluation Figure 3 shows some typical rules: 1,2,3 are good, 4,5 are fair, and 6 is bad. The main problems of these rules include: The extracted events are not incomplete, which makes the rules less informative, such as rule 4 and 5. The causality between cause event and effect event is not very strong, which should be attributed to the design of causal patterns and the process of rule induction, such as 4 and 6. Some other problems also exist, such as verb disambiguation when normalizing predicates, noun disambiguation when generalizing rule instances.

1 (Y,'价格/price','下降/fall','',''):-(X,'产量/yield','过剩/surplus','',''),isA(X,'燃料/fuel'),isA(Y,'自然资源/natural resource'),madeof(X,Y)conf:0.566
2 (Y,'价格/price','下降/fall','',''):-(X,'罢工/strike','结束/stop','',''),isA(X,'国家/nation'),isA(Y,'金属/metal'),atlocation(Y,X)conf:0.582
3 (X,'面积/area','减少/fall','',''):-(X,'价格/price','下降/fall','',''),isA(X,'作物/crop')conf:0.588
4 (X,'增长率/growth rate','下降/fall','',''):-(X,'储蓄率/saving rate','下降/fall','',''),isA(X,'国家/nation')conf:0.719
5 ('',Y,'适合/fit','',''):-('',X,'下降/fall','',''),isA(X,'产品/product'),isA(Y,'作物/crop'),madeof(X,Y)conf:0.563
6 ('','','增加/increase',Z,'销量/sales'):-('',X,'减少/fall',Y,'依赖性/dependence'),isA(X,'国家/nation'),isA(Y,'自然资源/natural resource'),isA(Z,'燃料/fuel'),madeof(Z,Y)conf:0.572

Figure 3: Examples of Typical Rules

[2] http://www.swi-prolog.org/
[3] http://finance.sina.com.cn

Event Graph With these rules, we deduce many rule instances with Prolog and pick out a tiny subgraph about rise and fall events, in Figure4, to show the power of the rules.

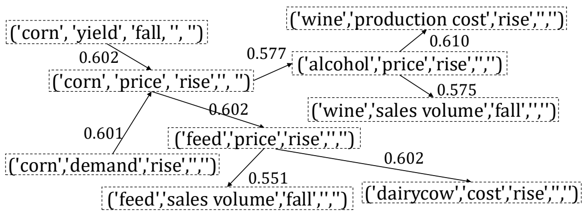

Figure 4: Rule Deduction. As space is limited, we only show the English version and omit the rules used in the reasoning process.

3.3 Comparison with existing Knowledge Bases

We compare our rules with causal part of other knowledge bases in various aspects in Table 1. We can see our causal knowledge representation is more expressive and informative, and the automatic knowledge acquisition is very convenient.

3.4 Ablation Study

In this section, we explore the contributions of the various components of our rule learning framework.

Causal patterns statistic The matched sentences distribution over 3 groups of patterns is shown in Table 2. All patterns in one group have different causal cue words literally but the same meaning. It shows the third pattern group is more rigorous than the first two groups but has lower usage. Probably because more logical thinking is needed when editing news using more rigorous patterns.

External Knowledge Bases The following is some statistics of external knowledge bases used in the rule learning framework. The size of the lexicon is 12,624, obtained from 'Industrial classification for national economic activities'[4], which determines which event role in the rule instance can be generalized. To our knowledge, most existing Chinese taxonomic knowledge bases, such as CN-Probase[**?**], zhishi.me[**?**], are constructed from online-encyclopedia, which suffer that the concepts inside are far less than Probase and they have no probabilistic character. So we translate Probase to get 11,292,493 Chinese 'IsA' pairs. To our knowledge, there exists no large-scale Chinese commonsense knowledge base, so we translate the English part of ConceptNet and merge the Chinese part to get 2,085,681 Chinese triples. We randomly sample 500 items from translated Probase and ConceptNet, respectively, and the accuracies after the human evaluation are **87.8%**(close to the accuracy of original Probase 92.6%) and **91.6%**.

Open Event Extraction Since our event structure scheme is plain and straightforward, we choose the reliable Stanford CoreNLP tool to extract the rule instances. The number of rule instances extracted after rule instance distilling submodule is 7,835,403. Since most of them are discarded in the learning process, the number of rule instances really used for rule induction is 78,098 with an accuracy of **48.5%** (we also sample 200 rule instances and manually evaluate them).

To sum up, our framework is a pipeline undergoing rule instance extraction(accuracy 48.5%), constrain relations addition(accuracy of ConceptNet 91.6%), and rule induction(accuracy of Probase 87.8%). Thus, the accuracy can only reach **39.0%** (48.5%*91.6%*87.8%) at the maximum, which is close to our evaluation(32.5%) of the final rules.

3.5 Application: Futures Price Prediction

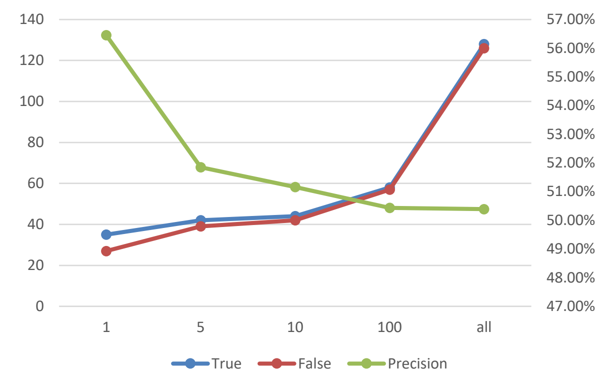

Figure 5: Futures Price Prediction.

We choose futures price prediction because the futures are common and concrete things existed in ConceptNet and Probase, such as corn, oil, etc. We follow similar experimental settings in [**?**]. From 2018/1/1 to 2018/11/2, we collect all the headlines and the price change of 15 futures as test data, which include **851** price change events (The price change of more than 1% relative to the previous day is an event and we only focus on rise or fall events).

Baseline models: EB_CNN model [**?**], the state of the art model in stock price prediction, uses a deep convolutional neural network to model both short-term and long-term influences of events on stock price movements, and the accuracy of futures prediction is **54.2%**. Other models in [**?**], such as EB_NN, WB_CNN, and WB_NN can achieve **53.0%**, **53.2%**, and **53.5%**, respectively. These accuracies of futures prediction are lower than the accuracies of stock prediction shown in the paper. It may be because the factors affecting the futures price are far less than the stock price and the futures price is much more stable than the stock price, which makes useful training information about the futures less and further affects the accuracy of the models.

Our approach: For each actual future price change event , we get the news headlines for the previous month before this event. For each news headline, we extract the event, use Prolog to reason based on the rules and external knowledge bases, and get the top K inferred events sorted by the confidence. We may have m*K inferred events for this event, m is the number of events occurred in this month. Here, we select the price change events(rise or fall) of the future in this actual future price change event from m*K events and calculate the weighted sum of their confidences(rise event weights 1 and fall event weights -1). If the sum value is positive, we predict

[4]http://www.stats.gov.cn/Tjsj/tjbz/hyflbz/

Name	Number	Domain	Unit	Data Structure	Information	Source	Precision
CausalNet	**62,675,002**	**Open**	word	(-)	rich	**automatic**	-
(‘drink’,‘accident’,36)							
ConceptNet	89,416	**Open**	short text	unstructured	rich	crowdsourcing	**100%**
(‘smoking’,‘/r/Causes’,‘cancer’)							
FrameNet	59	**Open**	frame	**structured**	richer	crowdsourcing	**100%**
Killing(Killer,Place,Means,Victim,Instrument),CausativeOf,Death(Protagonist,Place,Manner,Time)							
ATOMIC	568,312	**Open**	**logic event**	semi-structured	much richer	crowdsourcing	86.2%
If “PersonX pays PersonY a compliment”, Then “PersonY will smile”							
Ours	50,000	Finance	**logic event**	**structured**	**richest**	**automatic**	32.5%
(Z,‘price’,‘rise’,‘’,‘’):-(‘’,X,‘suffer’,Y,‘attack’),isA(X,‘country’),isA(Y,‘disaster’),isA(Z,‘metal’),atLocation(Z,X) conf:0.842							

Table 1: Comparison with existing knowledge bases

Pattern template	Priority	Number	Rate
因为 A,B	1	2000242	48.32%
A,所以 B	2	1530311	36.96%
因为 A, 所以B	3	576851	14.72%

Table 2: Number of sentences extracted by causal patterns. A is a cause span and B is an effect span. Word ‘因为’ represents a group works like 由于,是因为,因为,缘于,归因于,原因是,鉴于, and word ‘所以’ represents a group of words like 所以,因而,因此,故而,因故,导致,招致,以致,引致,诱致,致使,造成,使得,从而使,于是,为此

this future price as a rise event, otherwise as a fall event. If get no related events changing the future’s price, do not make prediction. We compare this prediction with the actual price change to evaluate the reasoning effect. Figure 5 shows the average prediction result. It shows the more predicted events inferred from the Prolog(by increasing K) we use, the lower the prediction accuracy is(from **56.5%** to **50.4%**), and the more futures events we can predict(from **62** to **254**).

To sum up, our rule-based prediction approach can have a higher prediction accuracy (56.45%) and better interpretation ability with a low recall rate, which is very practical in life.

3.6 Downloading and Demo

The translated Chinese Probase and ConceptNet and learned rules are available at URL. We built a demo to demonstrate the reasoning process at URL. We also developed an application demo of futures prices change triggering that can monitor news from around the world in real time, find the news that may cause futures prices changes, and alert users. Visit URL.

4 Related Work

Knowledge Representation and Reasoning Neural-based knowledge representation, such as [**?**] and [**?**], which uses the ‘translation’ method to embed knowledge graph into dense vectors, is effective for many downstream Natural Language Processing(NLP) works [**?**]. However, it has problems of uninterpretability and weak reasoning ability. In addition, there are many symbol-based knowledge representations. For example, ConceptNet [**?**] uses the relation ‘/r/Cause’ to represent causal knowledge, such as (‘smoking’, ‘/r/Causes’,‘cancer’). It suffers that such specific representation makes the knowledge less informative and less expressive. CausalNet[**?**] also represents the event with a word. FrameNet[**?**] uses ‘Causative_of’ to connect causal frames, which are too rough to clearly express causal knowledge. The semi-structured causal knowledge representation in ATOMIC [**?**] is also unfriendly for machines to reason.

Rule Acquisition Different knowledge representations have different knowledge acquisition methods. Here, we mainly review how to acquire knowledge of logic rule. The Inductive Logic Programming(ILP)[**?**; **?**; **?**] makes strong assumptions, such as high-quality training data, closed-world assumption, and so on, which are inappropriate to handle the extracted data from Web text. Therefore, SHERLOCK system [**?**] learns the first-order horn clauses in a top-down manner, which first identifies the classes(concepts in our rules) and relations(predicates in our events), enumerates all their combinations, and keep good ones as final rules. However, due to a large number of predicates and concepts in large web texts, as well as the complexity of the rule structure, their combination will be explosive in our case. So, we propose a bottom-up framework based on MDL to learn rules.

Financial Market Price Prediction There are many works on financial market price forecasting using text information, mainly on stock prices. [**?**] attempts to use structured event information to predict stock prices. Furthermore, [**?**] enriches event representation with the knowledge base to improve prediction accuracy.

5 Conclusion and Outlook

In this paper, we design a novel and powerful causal knowledge representation scheme based on the logic rule with the ability of uncertainty reasoning and we propose a rule learning framework for obtaining rules from large unstructured text. The experiments show that the rules learned are reasonable and effective. In the future, we would like to improve the main components in the rule learning framework, such as rule instance extraction, external knowledge bases, and rule induction. Besides, we will also explore other downstream applications based on this reasoning system, such as consumption intent prediction, stock price prediction, and reasoning-based information retrieval.

References

[Bach *et al.*, 2017] Stephen H. Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. Hinge-loss Markov random fields and probabilistic soft logic. *Journal of Machine Learning Research (JMLR)*, 2017.

[Baker *et al.*, 1998] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. *Proceedings of the 36th annual meeting on Association for Computational Linguistics* -, 1998.

[Bergadano and Gunetti, 1996] Francesco Bergadano and Daniele Gunetti. Inductive Logic Programming. 1996.

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, 2013.

[Chang and Choi, 2004] Du-Seong Chang and Key-Sun Choi. Causal relation extraction using cue phrase and lexical pair probabilities. In *International Conference on Natural Language Processing*, 2004.

[Cui *et al.*, 2016] Wanyun Cui, Xiyou Zhou, Hangyu Lin, Yanghua Xiao, Haixun Wang, and Wei Wang. Verb Pattern : A Probabilistic Semantic Representation on Verbs. *Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016)*, 2016.

[De Raedt *et al.*, 2007] Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. Problog: A probabilistic prolog and its application in link discovery. In *IJCAI*, 2007.

[Ding *et al.*, 2015] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Deep learning for event-driven stock prediction. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[Ding *et al.*, 2016] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Knowledge-Driven Event Embedding for Stock Prediction. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016.

[Li *et al.*, 2016] Xiang Li, Aynaz Taheri, Lifu Tu, and Kevin Gimpel. Commonsense knowledge base completion. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016.

[Luo *et al.*, 2016] Zhiyi Luo, Yuchen Sha, Kenny Q Zhu, Seung-won Hwang, and Zhongyuan Wang. Commonsense Causal Reasoning between Short Texts. *Proceedings, Fifteenth International Conference on Principles of Knowledge Representation and Reasoning (KR 2016)*, 2016.

[Manning *et al.*, 2014] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 2014.

[Muggleton, 1996] Stephen Muggleton. Learning from positive data. In *International Conference on Inductive Logic Programming*, 1996.

[Niu *et al.*, 2011] Xing Niu, Xinruo Sun, Haofen Wang, Shu Rong, Guilin Qi, and Yong Yu. Zhishi. me-weaving Chinese linking open data. *The Semantic Web–ISWC ...*, 2011.

[Pearl and others, 2009] Judea Pearl et al. Causal inference in statistics: An overview. *Statistics surveys*, 2009.

[Quinlan, 1990] J. R. Quinlan. Learning Logical Definitions from Relations. *Machine Learning*, 1990.

[Sap *et al.*, 2018] Maarten Sap, Ronan LeBras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. Atomic: An atlas of machine commonsense for if-then reasoning. *Proceedings of the 33th Conference on Artificial Intelligence (AAAI 2019)*, 2018.

[Schoenmackers *et al.*, 2010] Stefan (University of Washington/Turing Center) Schoenmackers, Oren (University of Washington/Turing Center) Etzioni, Daniel S. (University of Washington/Turing Center) Weld, and Jesse (Katholieke Universiteit Leuven) Davis. Learning first-order horn clauses from web text. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 2010.

[Speer *et al.*, 2016] Robert Speer, Joshua Chin, and Catherine Havasi. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. 2016.

[Waldmann and Hagmayer, 2013] Michael R Waldmann and York Hagmayer. Causal reasoning. 2013.

[Wang *et al.*, 2017] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2017.

[Wu *et al.*, 2012] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q. Zhu. Probase: A probabilistic taxonomy for text understanding. *Proceedings of the 2012 ACM SIGMOD ...*, 2012.

[Xu *et al.*, 2017] Bo Xu, Yong Xu, Jiaqing Liang, Chenhao Xie, Bin Liang, Wanyun Cui, and Yanghua Xiao. Cn-dbpedia: a never-ending chinese knowledge extraction system. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 2017.

[Zhao *et al.*, 2017] Sendong Zhao, Quan Wang, Sean Massung, Bing Qin, Ting Liu, Bin Wang, and ChengXiang Zhai. Constructing and Embedding Abstract Event Causality Networks from Text Snippets. *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining - WSDM '17*, 2017.

[Zhu *et al.*, 2010] Kenny Q Zhu, Kathleen Fisher, and David Walker. Incremental learning of system log formats. *ACM SIGOPS Operating Systems Review*, 2010.