Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

摘要—Causality 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. Much interesting information appears as natural language text, which must be processed and analyzed to derive valuable knowledge. This project aims to build a financial causality knowledge base by analyzing a large online financial text trying to capture the causal strength of different finance-related events. The rules in this knowledge base can be used to predict financial events and generate alerts in financial trading.

***Index Terms*—rule, causality, Knowledge base**

I. INTRODUCTION

This document is a model and instructions for LaTeX. Please observe the conference page limits. 举例 ConceptNet5 and WebBrain and WeChild 其中的关系

II. PROBLEM DEFINITION

In this section, we will present the definitions of all necessary concepts used through the paper, and formally state the focal problem to be solved.The list of major symbols and notations in this paper is summarized in the following table.

Identify applicable funding agency here. If none, delete this.

表 I
TABLE OF NOTATIONS

Notation	Definition
rule instance	structured causality proposition pair
concept constraint	concept of a argument
relation constraint	relation constraint of two arguments
RS_{ij}	relation saliency of $i-th$ Relation for $j-th$ Rule
E_{c_i}	the instances of $i-th$ concept
rule	generalized rule instance with concept constraints and relation

介绍 knoeledge base 的样子

We represent the structure of such a cause proposition with SPO approach, But which is coarse to refine the information in such noisy texts. For example, We often extract the SPO triples like(' 价格', ' 上涨',''). So we expand SPO to NSPNO with the form of quintuple (Compound Noun Of Subj, Subj, Predicate, Compound Noun Of Obj, Obj). Then we can get (' 玉米',' 价格',' 上涨','','') which is much more specific than before, meanswhile we also consider the neg relation of the predicate, we mark it with the same position of predicate when extraction. After designing the proposition representation scheme, hopefully, we wish extract the rule instances like (Compound Noun Of Subj, Subj, Predicate, Compound Noun Of Obj, Obj)->(Compound Noun Of Subj, Subj, Predicate, Compound Noun Of Obj, Obj).

Problem Statement. Mining the causality rule hidden in unstructured text.

III. Approach

A. Overview

图 1. Overvie of the framework.

“Fig. 2” describes the whole picture of our framework. generally, It consists three modules: rule instances extraction, rules generalization and rules distilment. Rule instances extraction submodule mainly extracts the structured causality proposition pairs (also called rule instance) from liberal text via curated patterns. Rules generalization submodule generalizes the rule instances to capture more flex expression via the newly built knowledge base, based on current knowledge bases,such as Probase [1], ConceptNet5 [1], WebBrain [1] and WebChild [1]. Rules distilment submodule mainly focus on concept selection, relation selection, rule filtering, and rule ranking.

B. Rule Instances Extraction

Causality can be identified by the causal cues [1], Consider the following sentences:

1) 但由于国际石油价格一路攀高一直到现在创纪录的每桶五十几美元，导致目前合成橡胶价格快速上升，已经超过天然橡胶，像丁苯橡胶已从年初的每吨 11000 元上涨到现在的 14000 元。
2) 2005 年我国乙醇汽油消费玉米 1300 万吨，预计 2006 年消费玉米 1600 万吨，由于石油涨价，乙醇转化用粮需求将会快速增长，从而拉动玉米价格上涨。

Then we parse the sentence to extract the structured casuality proposition pairs. Here, We use the Stanford CoreNLP’s [1] dependency parser. We input the whole sentence, then we join the tokens with padding spaces between each two tokens. for sentence (1), It would be ” 但由于国际石油价格一路攀高一直到现在创纪录的每桶五十几美元，导致目前合成橡胶价格快速上升，已经超过天然橡胶，像丁苯橡胶已从年初的每吨 11000 元上涨到现在的 14000 元。”, Then we exploit the designed pattern to match the cause span and effect span. The causal cues(casual patterns) are shown in Table 1. We assign each pattern a priority, the higher the earlier it matches. In sentence (1), we can match the cause span ” 国际石油价格一路攀高一直到现在创纪录的每桶五十几美元” and the effect span ” 目前合成橡胶价格快速上升，已经超过天然橡胶，像丁苯橡胶已从年初的每吨 11000 元上涨到现在的 14000 元。”. In sentence (2), we can match the cause span ” 石油涨价，乙醇转化用粮需求将会快速增长” and the effect span ” 拉动玉米价格上涨。” We firstly parse the sentence, not match with patterns, Because we want to alleviate the parsing error with more context information fed into the parser.

After the cause tokens span and effect tokens span, We further extract the causality proposition pairs. We regard every tokens with each tokens spans with verb pos-of-speech as predicates, but we can get subject or object beyond the spans, which is another advantage of parsing the whole sentence instead of parse the cause span and effect span respectively. We also get the Compound nouns of subject and object with the dependency relation ’compound:nn’, and these nouns are write as the order appeared in the origin sentence. Sine maybe more than one predicates are in cause tokens span or effect tokens span, which means we can get more than one quintuple from cause tokens span or effect tokens span, we do cartesian product to get more pairs, and later, we will filter the low frequency pair.

表 II

CAUSAL CUES. A IS A CAUSE TOKENS SPAN, AND B IS AN EFFECT TOKENS SPAN.

pattern	priority
A, 所以 B	0
由于 A, B	0
A 导致 B	3
因为 A, 所以 B	3
由于 A, 导致 B	3
由于 A, 从而 B	3

C. Rules Generalization

After we get many rule instances from rule instances extraction submodule, Then we need generalize these rule instances, Since we want to more general rule to capture more general information: for example we hopefully generalize (' 国际石油', ' 价格', ' 攀高 @ 攀高', ", ") −− > (' 橡胶', ' 价格', ' 上升 @ 升高', ", ") to('X0', ' 价格', ' 攀高', ", ") −− > ('X1', ' 价格', ' 升高', ", ") where 'X0' IsA' 化石燃料' 'X1' IsA ' 产品' also we want to know the relation between X0,X1 is X0 'madeof' X1.

1) Construct Knowledge Base: Such a generalization operation mainly has two parts, one is finding Concept Constraint of Arguments and the other is finding Relation Constraint Between Arguments, So We need exploit external knowledge base.

Taxonomy Finding Concept Constraint of Arguments needs a taxonomy. Taxonomy is a IsA relation instance base. since chinese IsA relation items are too less, We have to enlarge it with english, and translate it. we exploit the items in Probase which are all IsA relation, the items with IsA relation in ConceptNet5,the items with IsA relation in Webbrain.

commonsense relation Finding Relation Constraint Between Arguments needs a facts base, intuitively, We find most the relation between two entities in cause and effect can be connected using commonsense. emiting the commnosense relation gap from cause to effect do not affect people's mind, But If we wan't to capture the general rule behind the rule instance, We need dig it out. We inevitably need a large commonsense base. So we build a new commonsense based on ConceptNet5, Probase, WebBrain, WebChild.

one pair of entity may have more than one relation. fot example a pair of (sweet corn,corn) has the relations /r/RelatedTo and /r/PartOf, obviously, /r/RelatedTo is less imformative than /r/PartOf, We design a algorithm to eliminate the relations which are semantically repeated.

Algorithm 1 Build Commonsense Relations

Require: all triples (arg1, rel, arg2) in CoceptNet5, WebBrain, WebChild

standardise the relations, existed in ConceptNet5 and WebBrain and WeChild.
initialize relationArguments= defaultdict(set), argumentsRelations=defaultdict(set)
for all $(arg1, rel, arg2)$ in trples **do**
  relationArguments[$(arg1, arg2)$].add(rel)
  argumentsRelaions[rel].add($(arg1, arg2)$)
end for
relationSet=relationArguments.keys()
relationPriority=Set()
for all (rel_1, rel_2) in combination(relaionSet) **do**
  if $Is_Relavant(rel_1, rel_2)$ **then**
    $p(rel_1|rel_2) = \frac{|E_{rel_1} \cap E_{rel_2}|}{|E_{rel_1}|}$
    $p(rel_2|rel_1) = \frac{|E_{rel_1} \cap E_{rel_2}|}{|E_{rel_2}|}$
    if $p(rel_1|rel_2) > p(rel_2|rel_1)$ **then**
      relationPriority.add((rel_1, rel_2))
    else
      relationPriority.add((rel_2, rel_1))
    end if
  end if
end for
for all $(arg1, arg2, rels)$ in argumentsRelaions **do**
  Remove the low priority of each two relations existed in rels.R
end for

2) Concept Constraint of Arguments: After getting the taxonomy, We can conceptualize the arguments of the predicates in cause proposition and effect proposition. Here we look up the taxonomy of each argument to give it a concept constraint. For one argument, We may find many concepts, But will choose a suitable one, later.

3) Relation Constraint Between Arguments: Conceptualizing one argument to a concept make the rule much more general, (’ 国际石油’, ’ 价格’, ’ 攀高 @ 攀高’, ”, ”) −− > (’ 橡胶’, ’ 价格’, ’ 上升 @ 升高’, ”, ”) to(’X0’, ’ 价格’, ’ 攀高’, ”, ”) −− > (’X1’, ’ 价格’, ’ 升高’, ”, ”) where ’X0’ IsA’ 化石燃料’ ’X1’ IsA ’ 产品’, For instance, coal IsA ’ 化石燃料’,phone IsA ’ 产品’, We instance X0 and X1 respectively with coal and phone. Then the rule means raising the price of coal would lead to the rubber’s rising, Which is obviously unreasonable. So we hopefully find the relation constraint between X0 and X1, and mostly the relations are commonsense, which is the intuition We build our knowledge base.

We find all the one hop relations between each two arguments existed in cause and effect proposition from our built knowledge base, which may consist many pair of argument, and many relations of one pair arguments, But we will choose suitable relation constraints later.

4) Predicate Normalization: Different literal predicate may expressive the same meaning. ”raise, rise, soar, increase, gain, enhance” have the same meaning, So we use the Ciling [1] to canonicalize the predicates. Ciling is a the largest word-level chinese resource, which split the chinese words into group according to the distance and relevance of word meaning. for each meaning group, we choose most frequent one appeared in text to represent the whole words in its group.

D. Rules Distilment

Argument Concept Selection. For one argument It may have several concepts return back from the taxonomy,So we need choose a proper concept.

Given $IsA(X, c_1), IsA(X, c_2), IsA(X, c_3)$, Return $IsA(X, c) c \in c_1, c_2, c_3$

1) **concept no-overlap.**

   if $IsA(c_1, c_2)$ remove c_2 since c_2 is too general calculate:

$$Overlap(c_1, c_2) = \frac{|E_{c_1} \cap E_{c_2}|}{min\{|E_{c_1}|, |E_{c_2}|\}} \tag{1}$$

   if $Overlap(c_1, c_2) > \theta$ remove c_i $wherei = argmax|E_{c_i}|$

2) **concept Relevance.**

$$CR(v, c_j) = p(v, c_j) log \frac{p(v, c_j)}{p(v)p(c_j)} \tag{2}$$

$$IsA(X, c_j) wherej = argmax_j CR(v, c_j)$$

Argument Relation Selection.

$$RS_{ij} = \frac{Count(rel_i, R_j)}{Count(R_j)} \tag{3}$$

pick rel_i for R_j as relation constraint, where $i = argmax_i RS_{ij}$

Rules Ranking.

IV. EXPERIMENT

In this section, we conduct extensive experiments to evaluate the effect of each submodule based on a real dataset crawled from financial news website, which contains 4,991,000 articles. Since most sentence has no casual cues and parsing a time-consuming process, We firstly filter out the sentences without casual patterns. all the algorithm in the following experiments are implemented in Python and run on a Intel Xeon 32 CPU(2.60GHz) and 173GB memory.

1) Rule Instances Extraction: **Pattern statistic** The ?? casual cues we used can be grouped into ?? sets, each containing cues of the same meaning but different form. Causal evidence distribution over these pattern sets is shown in “Fig. 2”

图 2. Number of sentence matched by cues

2) Rules Generalization: **our newly built Knowledge Base.** consistency rate: our new built contains ??? pairs of arguments , is 3/4 times of pairs in ConceptNet5.

3) Rules Distilment: we eveluate the ranked rules and 50% is good.

V. Error Analysis

A. Pattern Extraction

When use pattern, since the sentence is very noisy to extract the cause and effect which is hard to extract just depend on patterns. For Example, ” f”.

B. Rule Instance Extraction

The parser itself exist error for example ”” The causality relation is

C. Rule Generalization

Concept error. Predicate error. Relation error.

D. Rule Distilment

VI. Related Work

VII. Conclusion

VIII. Ease of Use

A. Maintaining the Integrity of the Specifications

The IEEEtran class file is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

IX. Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections IX-A–IX-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads—LaTeX will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “$\mathrm{Wb/m^2}$” or “webers per square meter”, not “webers/$\mathrm{m^2}$”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “$\mathrm{cm^3}$”, not “cc”.)

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \tag{4}$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(4)”, not “Eq. (4)” or “equation (4)”, except at the beginning of a sentence: “Equation (4) is . . .”

D. LaTeX-Specific Advice

Please use "soft" (e.g., `\eqref{Eq}`) cross references instead of "hard" references (e.g., `(1)`). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don't use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in LaTeX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you've discovered a new method of counting.

BibTeX does not work by magic. It doesn't get the bibliographic data from thin air but from .bib files. If you use BibTeX to produce a bibliography you must send the .bib files.

LaTeX can't read your mind. If you assign the same label to a subsubsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

LaTeX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it's supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won't be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively".
- In your paper title, if the words "that uses" can accurately replace the word "using", capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect", "complement" and "compliment", "discreet" and "discrete", "principal" and "principle".
- Do not confuse "imply" and "infer".
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al.".
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [7].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation.

Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. *Identify the Headings*

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. *Figures and Tables*

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 3”, even at the beginning of a sentence.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

表 III

Table Type Styles

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

图 3. Example of a figure caption.

Acknowledgment

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks . . .”. Instead, try “R. B. G. thanks. . .”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first . . .”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

参考文献

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.