Acquiring Bayesian Networks from Text

Olivia Sanchez-Graillet and Massimo Poesio

University of Essex
Department of Computer Science
Wivenhoe Park, Colchester CO4 3SQ, United Kingdom.
{osanch, poesio}@essex.ac.uk

Abstract

Causal inference is one of the most fundamental reasoning processes and one that is essential for question-answering as well as more general AI applications such as decision-making and diagnosis. Bayesian Networks are a popular formalism for encoding (probabilistic) causal knowledge that allows for inference. We developed a system for acquiring causal knowledge from text. Our system identifies sentences that specify causal relations and extracts from them causal patterns, taking into account connectives such as conjunction, disjunction and negation, and recognising causes and effects by analysing terms. The dependencies among the causes and effects found in text can be encoded as Bayesian networks. We evaluated our work by comparing the network structures obtained by our system with the ones created by a human evaluator.

Introduction and Motivations

Causal inference is one of the most fundamental reasoning processes (Glymour, 2003; Pazzani, 1991; Trabasso's paper in Goldman et al, 1999) and one which is essential for question-answering as well as more general AI applications such as decision-making and diagnosis. Methods for acquiring knowledge about causal rules are a prerequisite for the development of systems capable of causal inference in these applications, especially in complex domains (Girju, 2003; Kontos et al, 2002).

Bayesian Networks (Pearl, 1998) are a popular formalism for encoding probabilistic causal knowledge and for causal inference. Such networks are typically acquired from data (Mani & Cooper, 2001), but text is a rich source of information about causal relations that can be exploited, even though there are a number of problems to take into account (Hearst, 1999). In this paper we discuss domain-independent methods for acquiring from text causal knowledge encoded as Bayesian networks.

Background - Bayesian Networks

A Bayesian network (Pearl, 2000) is a directed acyclic graph whose arcs denote a direct causal influence between parent nodes (causes) and children nodes (effects). The nodes can be used to encode any random variable. For example, a person can be ill or well; the car engine can be working normally or having problems, etc. Such graph is associated with a probability distribution that satisfies the Markov Assumption. By using Bayesian networks it is possible to handle incomplete knowledge as well as to make predictions by using the conditional probability distribution tables (CPT). There is one table for each node, which describes the conditional probability of that node given the different values of its parents (Friedman & Goldszmidt, 1996). A disadvantage of these tables is that they can be huge because the size of the table is locally exponential to the number of parents of the node.

The complete joint probability distribution for the network is expressed by the CPTs for all the variables together with the conditional independences described by the network (Mitchell, 1997).

Identifying Causal Relations

Acquiring causal knowledge from text requires, first of all, identifying portions of text that specify a causal relation (henceforth **causal patterns**) between causes and effects (henceforth **events**) such as: "*Corruption and insecurity cause social problems*", "*Disease provokes pain or death*", "*Earthquake generates victims*" (Girju & Moldovan, 2002; Wolff et al, 2002); and second, analysing these causal patterns (a) taking into account the possible presence of connectives such as conjunction, disjunction and negation and (b) identifying causes and effects by analysing terms. These analysis steps are seldom discussed in the literature and have been the focus of our research. We consider each step in turn.

Finding causal patterns

Causal patterns can be expressed by cues such as connectives, as in "*the manager fired John **because** he was lazy*"; verbs, as in "*smoking **causes** cancer*"; or NPs, as in "*Viruses are **the cause of** neurological diseases*". After a preliminary analysis, we decided to concentrate in this first stage on causal patterns in which both events are expressed as noun phrases (ignoring cases such as in "*the manager fired John **because** he was lazy*"). We also decided to restrict the number of cues to the cause words in Roget Thesaurus found to be the most frequent in texts using Google, together with the causal verbs proposed by Girju and Moldovan (2002). Girju and Moldovan focused on explicit intra-sentential syntactic patterns of the forms <NP1 verb NP2> and <NP1 cause_vb NP2>. In the latter they use WordNet (Fellbaum, 1998) causal relations to find noun concepts of the verbs with nominalizations. They developed a method for automatic detection of causation patterns and semi-automatic validation of ambiguous lexico-syntactic patterns that refer to causal relationships. In this work we used the causal verbs that

they found to be the most frequent and less ambiguous such as *lead (to), derive (from), result (from)*, etc.

Connectors that denote implicit causal relationships like *when, after* and *with* identified by Khoo (Khoo et al, 2000) were not considered in this work since deeper semantic analysis is needed.

Examples of causal patterns identified by our system include: "*Anemia are **caused by** excessive hemolysis*", "*Hemolysis is a **result of** intrinsic red cell defects*", and "*Splenic sequestration **produces** anemia*".

Analysing causal patterns I – Connectives

The first step of analysis of the causal patterns deals with connectives (Jaegwon, 1971; Rader & Sloutsky, 2001; Cheng & Novick, 1991). Our system can deal with text in which events are conjoined or disjoined, as in "*Corruption, pollution **and** insecurity cause social problems*" and "*Bacteria, germs **or** virus provoke diseases*". The system also detects negated causal patterns, as in "*Victims were **not** caused by the earthquake*", and ignores them.

Rader & Sloutsky (2001) argued that conjunctions are better viewed as unit causes/effects, whereas disjunctions and conjunctions should be decomposed. As a result, our system treats a conjunction like "*Corruption and insecurity*" as a single event, whereas in the case of "*Bacteria, germs or virus*" three separate **atomic causal patterns** are identified, each of which contributes to the estimation of a separate conditional probability in the specification of the Bayesian network.

We use *and* and comma (,) for identifying conjunction and *or* and comma (,) for disjunction (without considering more complicated negative disjunctions like *nor* as in "*Neither men nor woman cause bad situations*"). For example, in "C*ity pollution, delinquency and crime are the result of poverty and growing*", "*poverty and growing*" is taken as a single atomic causal pattern. The same applies to "C*ity pollution, delinquency and crime*", which is taken as a single event. In "*Bacteria, virus or other microorganisms provoke diseases*" the relations obtained from splitting the phrase are A) "*Bacteria cause diseases*" B) "*Virus cause diseases*" and C) "*Other microorganisms cause diseases*".

In general, an event may generate a number of relations determined by the number of disjuncts contained both in the cause and in the effect.

Analysis II - Cause / Effect Generalisation

Generalisation is important for counting frequencies of events, and therefore for the probability calculation. We experimented with a further and optimal step, in which the system uses WordNet (Fellbaum, 1998) to normalize causes and effects by checking whether either has already been expressed in an alternative form using synonyms. For example, the system discovers that "*The centre of the mitochondria*" and "*The core of a cell microorganism*" express the same event. This process is explained in more detail in what follows.

The synonyms of the head nouns of the NPs that express events in causal patterns are obtained and compared. If their lemmas have synonyms in common they are considered similar.

Since the WordNet API does not provide a simple function for getting synonyms directly, the system obtains the synonyms of a word by getting the top-hierarchy level hypernyms (bold letters) of all the senses of this word's lemma, as in the following example.

```
Synonyms/Hypernyms (Ordered by Estimated Frequency) of
noun nutrition
3 senses of nutrition

Sense 1
nutrition
    => organic process, biological process
        => natural process, natural action, action, activity
            => process
                => phenomenon
Sense 2
nutriment, nourishment, nutrition, sustenance, aliment,
alimentation, victuals
    => food, nutrient
        => substance, matter
            => entity, physical thing
Sense 3
nutrition
    => science, scientific discipline
        => discipline, subject, subject area, subject field, field,
field of study, study, bailiwick, branch of knowledge
            => knowledge domain, knowledge base
                => content, cognitive content, mental object
                    => cognition, knowledge, noesis
                        => psychological feature
```

Figure 1 WordNet Synonyms/Hypernyms

Figure 1 shows the senses of *nutrition*, whose lemma is the same, as well as the synonyms/hypernyms of each sense obtained from WordNet. The system takes only the top level of hypernyms of each sense in order not to lose precision when considering lower levels. Thus, the set of synonyms obtained for *nutrition* is: [organic process, biological process, food, nutrient, science, scientific discipline].

The concern in this stage is to form interesting building blocks for collecting frequencies. Such blocks are formed by the head nouns of each NP. For example in (1) "*Excellent health is caused by protein absorption*" the cause's head noun is *absorption* and the effect's head noun is *Health*. Here, we have a compound nominal formed by two nouns. We take the final noun (*absorption*) as head noun.

In the sentence (2) "*Health is caused by good nutrition*" the cause's head noun is *nutrition* and the effect's head noun is *Health*. In this example, the set of synonyms of both head nouns are compared. If they have at least one common synonym, they are considered similar. The synonyms of *absorption* are [natural process, natural action, action, activity, social process, organic process, biological process, attention, cognitive state, state of mind]. In this case, *nutrition* and *absorption* have the common synonyms "*organic process*" and "*biological process*". Thus, these causes are considered equal. The system indicates that the cause of relation (1) is the same as the cause of relation (2), and also that the –cause,

effect- of relation (1) is similar to the –cause, effect- of relation (2). It means that both causal patterns are equal. The frequencies of the causes *nutrition* and *absorption* are incremented, and in the graph the cause node is labeled as "*nutrition / absorption*".

It should be clear from the example above that generalisation may lead to a loss of precision as illustrated by the following example: "*Love may cause either happiness or sadness*" we obtain A)"*Love cause happiness*" and B)"*Love cause sadness*". Both *happiness* and *sadness* are feelings, so if generalisation is performed, they are taken as similar terms. However, in this case these terms are actually antonyms. So the result is an incorrect event. Also, in this example we can observe that the word "*may*" denotes a certain degree of causality. However, in the current work the strength of causal relations has not been considered.

Other inaccuracies in generalisation may be caused by lexical ambiguity or by lemmatization problems. For these reasons, generalisation is only optional. Another problem when generalising, is that WordNet does not know some technical words like *hemosiderinuria*. Therefore it cannot return any synonym/hypernym of such word.

In order that the annotator can analyse the causal patterns obtained, the system displays them as well as the numbers of the patterns that have similar causes or effects.

Architecture

The system was developed in Java version 1.4.0, using the XML DOM model. It performs term generalisation as an optional choice, computes conditional probabilities for each node and generates an XML file that encodes the Bayesian network structure and the conditional probability tables. This file can also be saved as BIF (Bayesian Interchange Format) making possible to handle the network with different software that provide other tools for Bayesian networks such as the generation of cases and datasets. The connectors (like *caused by*) used by the system to identify the causal patterns are also stored in an XML file. Such file can be updated in order to delete, modify of include new connectors.

Our system takes as input a text tokenized, POS-tagged and partially parsed by using the LT-XML software developed by the University of Edinburgh's LTG group (http://www.ltg.ed.ac.uk/software/xml/index.html). (Lt Chunk is a partial parser that only recognises verbal expressions and noun expressions, but not prepositional phrases). For example, the sentence "*Bacteria, germs or virus cause diseases*" is first parsed and POS-tagged, producing:

```
<ne id="id1"><W pos="NN">bacteria</W></ne>
<W pos=",">,</W>
<ne id="id2"><W pos="NNS">germs</W></ne>
<W pos="CC">or</W>
<ne id="id3"><W pos="NN">virus</W></ne>
<ve id="id4"><W pos="VBP">cause</W></ve>
<ne id="id5"><W pos="NNS">diseases</W></ne>
```

Figure 2: Preprocessed text

After analysis, the system outputs one or more atomic causal patterns (ACP):

Cause	Connective	Effect
bacteria	provoke	disease
germens	provoke	disease
virus	provoke	disease

Figure 3: Example of a system output

The system uses the ACPs to estimate the conditional probabilities to be encoded in the distribution tables of the Bayesian network. It first estimates the conditional probability: P(effect | cause) by Maximum Likelihood Estimation (C(effect, cause) / C(cause)) computed by storing all the ACPs found in text and then counting the frequencies of events that are similar. These probabilities are then output in a format suitable for the Bayesian Network construction software CIspace V.2.5 (http://www.cs.ubc.ca/labs/lci/CIspace) that is needed to produce and analyse a Bayesian network, as shown in the following figure.
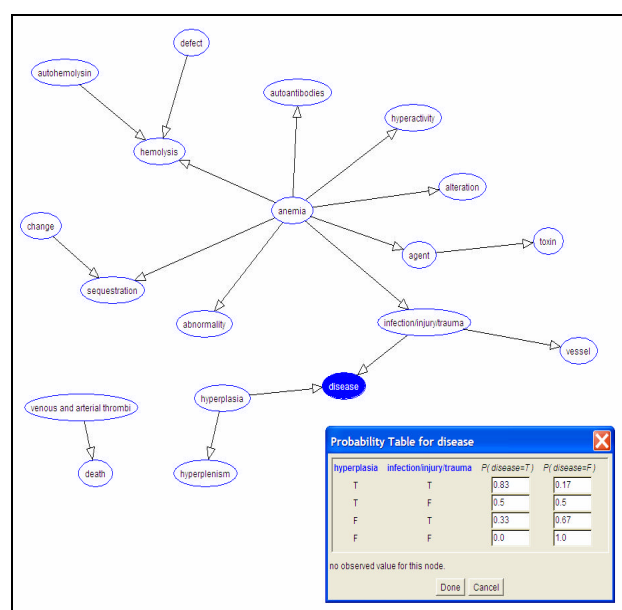

Image 1: Bayesian network obtained from text.

In the figure above, we can see that generalisation was performed. Thus, *infection, injury* and *trauma* were taken as the same event. In the same figure, the probability table shows the values that the variable *disease* can take for different combinations of the binary states of its parents. Such probabilities were calculated by the union of the conditional probabilities of the variable (effect) given its parents (causes).

By analysing the probabilities of each variable we can perform inference tasks. For example, if *hyperplasia* takes place, then it is more likely that a *disease* occurs than if *infection* is present. In addition, the probability values can be modified in order to predict what could happen in the

model obtained by making use of other source of knowledge apart from corpus.

We also can observe that the system only constructs network with binary variables and discrete domain.

Preliminary Evaluations

The system has been tested in texts from five different domains, including medical diagnosis, health care information, software failure diagnostics, engine tip forums and social forums, all of them obtained from the web. These texts are specific in their respecting topic.

We performed a subjective evaluation of the networks that simply consisted in the comparison between the structure of the network generated by the system and the structure of a network created manually by observing the causal patterns in text. The structure generated by the system was quite similar to the one created manually (same name of variables and same number of nodes and arcs indicating dependencies). Seven texts were analysed and in general they matched the structure of their corresponding reference network with a precision of 60%. However, precision can vary depending on the causal connectors used and if generalisation is carried out or not. For the current evaluation generalisation was performed and the connectors used in all networks were the same.

It is important to evaluate the degree of accuracy of the network with the help of a specialist in the text topic. This will be done in future work as well as a more accurate evaluation by measuring the distance between the distribution probabilities of a gold standard network and one obtained by our system.

Other factors that made the output network vary were the presence of bi-directional arcs in the network, due to the use of ambiguous causal connectors (as "*associated with*") that do not express a clear dependency direction, as well as the occurrence of events expressed as anaphoric expressions that produce ambiguous events.

Future Work

Future work will focus on medical domain since we found higher occurrence of causal patterns in it given that diseases can be diagnosed or cure by recognising their causes as well as the effects of prescriptions.

Moreover, future work will include improved evaluation methods and term extraction methods. As well as more focused evaluations, we plan to measure how well the network works when performing tasks such as obtaining accurate inferences, answering questions about the content of the text or supporting decision-making. For term identification, we plan to consider ontologies formed by modifiers and nouns, the recognition of specialised terms of the topic when generalisation takes place as well as the integration of anaphora resolution. Finally, we will consider the degree of causality encoded in the use of auxiliaries (as *may, could* and *must*) as well as adverbs (such as *strongly, slightly*) in order that the system gets more precise probabilities.

References

Cheng, P.W. & Novick, L. R. (1991). Causes versus enabling conditions. Cognition, 40, (pp. 83-120).

Friedman, N. & Goldszmidt, M. (1996). Learning Bayesian Networks with Local Structure. Proceedings of the Twelfth Conference for Uncertainty in Artificial Intelligence (UAI-96), (pp. 252-262). San Francisco, CA. Morgan Kaufmann Publishers.

Girju, R. & Moldovan, D. (2002). Text Mining for Causal Relations, FLAIRS Conference 2002, (pp. 360-364), Pensacola Beach, Florida, USA.

Girju, R. (2003). Automatic Detection of Causal Relations for Question Answering, The ACL 2003 Workshop on Multilingual Summarization and Question Answering, Sapporo, Japan.

Glymour, C. (2003). Learning, prediction and causal Bayes nets. TRENDS in Cognitive Science Vol.7 No.1.

Goldman, S.R., Graesser, A.C. & Broek, P.W. (1999). Narrative comprehension, causality, and coherence: Essays in Honor of Tom Trabasso, Mahwah, NJ: Erlbaum Associates.

Hearst, M. (1999). Untangling Text Data Mining, Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics.

Jaegwon, K. (1971). Causes and Events: Mackie on Causation. The Journal of Philosophy, Vol. 68, No. 14, (pp. 426-441).

Khoo, Ch., Chan, S. & Niu, Y. (2000). Extracting Causal Knowledge from a Medical Database Using Graphical Patterns, 38th Annual Meeting of the Association for Computational Linguistics (ACL), Hong Kong, China.

Kontos J., Malagardi I., Peros J. & Elmaoglou A. (2002). System Modeling by Computer using Biomedical Texts, Res-Systemica, Vol.2, Special Issue-Proceedings of the fifth European Systems Science Congress, Crete.

Mani, S. & Cooper, G. (2001). A Simulation Study of Three Related Causal Data Mining Algorithms. Proceedings of the International Workshop on Artificial Intelligence and Statistics, (pp73-80), Morgan Kaufmann, San Francisco, California.

Mitchell T. (1997). Machine learning, Published Boston, Mass: McGraw-Hill.

Rader, A. W. & Sloutsky, V. M. (2001). Conjunction bias in memory representations of logical connectives. Memory & Cognition, 29(6), (pp. 838-849).

Pazzani, M.J. (1991). A Computational Theory of Learning Causal Relationships, Cognitive Science, Vol. 15, No. 3, (pp. 401-424).

Pearl, J. (1998). Probabilistic reasoning in intelligent systems: networks of plausible inference, San Mateo, California: M. Kaufmann.

Pearl, J. (2000). Causality: models, reasoning, and inference, Cambridge Press.

Wolff, P., Song, G. & Driscoll, D. (2002). Models of Causation and Causal Verbs. In the 37th Meeting of the Chicago Linguistics Society, Main Session, Vol. 1 (pp. 607-622), Chicago Linguistics Society.