Efficient Context-Sensitive Plausible Inference for Information Disclosure

P.D. Bruza and L.C. van der Gaag

RUU-CS-93-42
December 1993

Utrecht University

Department of Computer Science

Padualaan 14, P.O. Box 80.089,
3508 TB Utrecht, The Netherlands,
Tel. : ... + 31 - 30 - 531454

ISSN: 0924–3275

2 Background and Preliminaries

The main features of the logic-based approach to information disclosure will be highlighted as it provides the integrating framework for the ideas presented in this paper.

2.1 The Language of Index Expressions

Consider a set $\mathcal{O}$ of *information objects.* To facilitate their disclosure, each object in $\mathcal{O}$ is characterized by a description of its contents. In this paper, it is assumed that these descriptions are expressed in the *language of index expressions.*

Informally speaking, an *index expression* consists of a number of *terms*, separated by means of *connectors* modelling the relationships between these terms. Terms are taken from a given set T of terms and correspond to nouns, noun-qualifying adjectives and noun phrases; connectors are taken from a set C of connectors and are basically restricted to the prepositions and the so-called *null connector.* More formally, the language $\mathcal{L}(T, C)$ of index expressions over T and C is defined by the following syntax:

$$
\begin{aligned}
\mathsf{Expr} &\rightarrow \epsilon \mid \mathsf{Nexpr} \\
\mathsf{Nexpr} &\rightarrow \mathsf{Term}\ \{\mathsf{Connector}\ \mathsf{Nexpr}\}^* \\
\mathsf{Term} &\rightarrow \mathsf{t},\ \mathsf{t} \in T \\
\mathsf{Connector} &\rightarrow \mathsf{c},\ \mathsf{c} \in C
\end{aligned}
$$

where ϵ denotes the empty index expression. Examples of index expressions are people in need of information and effective $\circ$ information $\circ$ retrieval; in the latter index expression $\circ$ denotes the null connector.

The description of an information object $O \in \mathcal{O}$ is drawn from the language of index expressions and is denoted as $\chi(O)$; an index expression in $\chi(O)$ is called an *axiom* of O. For details of how index expressions can be automatically derived from the titles of information objects, the reader is referred to [Bruza, 1993].

Building on the notion of index expression, we introduce the notion of *power index expression:* the power index expression of an index expression is the set of all its index subexpressions. More formally, the power index expression of an index expression i, denoted as $\mathcal{P}(i)$, is the set

$$\mathcal{P}(i) = \{j \mid j \subseteqq i\}$$

where $\subseteqq$ denotes the is-subexpression-of relation. The power index expression of a given index expression forms a lattice where the underlying ordering is $\subseteqq$; the top of this lattice is the index expression itself and the bottom is the empty index expression ϵ. As an example, Figure 1 depicts the power index expression of effective $\circ$ information $\circ$ retrieval.

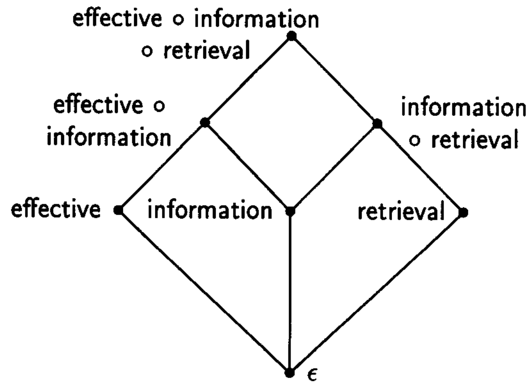

Figure 1: An Example Power Index Expression

For a set of information objects, a core set $\mathcal{I}$ of index expressions is generated. Each element from this set gives rise to a power index expression. These power index expressions may have some non-trivial overlap. Now, by forming the union of all these power index expressions, that is, by taking $\bigcup_{i \in \mathcal{I}} \mathcal{P}(i)$, a lattice-like structure is yielded. This structure will be termed a *lithoid.* As an example, Figure 2 shows the lithoid yielded by the power index expressions of the expressions effective $\circ$ information $\circ$ retrieval and people in need of information.

2.2 Strict Inference and Plausible Inference over Index Expressions

Information disclosure relative to a set of information objects $\mathcal{O}$ is driven by a request q of a searcher; the aim of the disclosure is to find all information objects $O \in \mathcal{O}$ that are relevant to this request. To this end, q is expressed in the language of index expressions $\mathcal{L}(T, C)$ just as the object characterizaions are, that is, $q \in \mathcal{L}(T, C)$. The problem of disclosure is approached by applying so-called *inference rules* for deriving the request from object descriptions. We distinguish between rules of *strict* inference and rules of *plausible* inference.

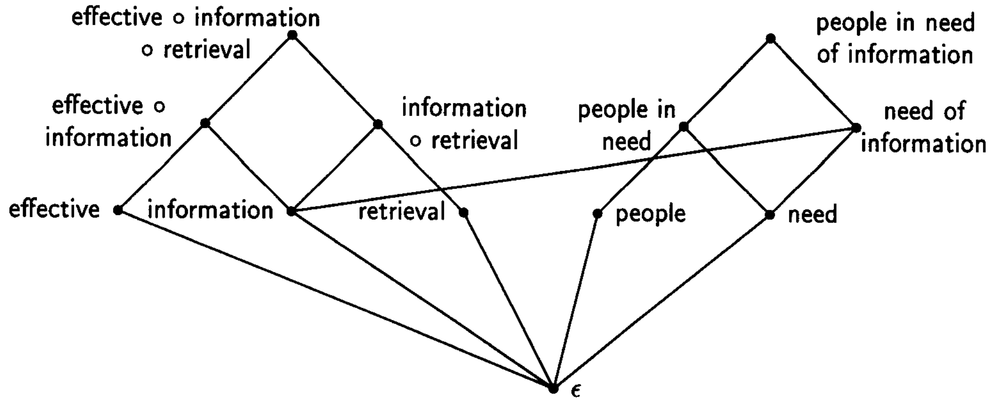

Figure 2: An Example Lithoid

Modus continens is a rule of strict inference; it may be looked upon as deduction by containment. Consider two index expressions i and j in $\mathcal{L}(T, C)$. Then, if j is an index subexpression of i, j can be derived from i through *modus continens*, denoted as

$$i \vdash_{MC} j$$

If a request q can be proven from the set of axioms of some object O by applying *modus continens*, then we are sure that O deals with, or *is about*, q. In that case, O is relevant with respect to q and should be returned in response. Note that the lithoid constructed from a core set of index expressions constitutes all index expressions derivable from this core by applying *modus continens*. For other rules of strict inference, refer to [Bruza, 1993].

If a request q cannot be (strictly) derived from the set of axioms of an information object O, this does *not* necessarily mean that O is not relevant to q: it only means that the axioms of O are too weak to establish relevance of O with respect to q. It will be evident that applying only rules of strict inference results in an imbalance between relevance and derivability. To alleviate this imbalance *plausible inference* is used. Plausible inference strives to generate high probabilities of relevance for those relevant information objects that escaped the strict inference mechanism. If for a given object the probability of relevance is high, then it might be returned in response to the request q after all.

Inference by refinement is a rule of plausible inference. Refinement of an index expression is making it more specific by adding a connector-term pair to it. As an example, consider the index expression information. This expression can be refined into the index expression need of information which can in turn be refined into people in need of information; these refinements result from the inverse $\subseteq$-relation over the language of index expressions and therefore are determined by the underlying lithoid.

Refinement can be taken as the basis for plausible inference: by applying refinement inference is rendered plausible in the sense that the derivation introduces index expressions that are no part of the original description of an object's contents and are not derivable through strict inference. Consider two index expressions i and j in $\mathcal{L}(T, C)$. Then, if i can be refined into j, j can be plausibly derived from i through *inference by refinement*, denoted as

$$i \mid\!\sim_{PR} j$$

Note that *inference by refinement* applies to a single index expression as is true for *modus continens*. From this observation, it follows that a derivation employing *modus continens* and *inference by refinement* takes the form of a sequence of transformations on an index expression transforming it into another one. An immediate consequence is that the relevance of an object with respect to a given request can be established by deriving the request from a single index expression in the object's description.

Modus continens and *inference by refinement* drive a *context-free* inference mechanism: in applying these rules, major parts of the context provided by an initial index expression from an object's description are discarded. It therefore is not possible to distinguish between the following two derivations:

$$\begin{array}{lll} \text{information} \circ \text{need} & \vdash_{MC} & \text{need} \\ & \mathrel{|\!\!\sim}_{PR} & \text{need of information} \end{array}$$

$$\begin{array}{lll} \text{information} \circ \text{need} & \vdash_{MC} & \text{need} \\ & \mathrel{|\!\!\sim}_{PR} & \text{need of food} \end{array}$$

because the initial context mentioning the need being an information need is discarded and therefore cannot be used further in the derivation. In the next section, we propose a context-sensitive rule of plausible inference over index expressions that is able to discriminate between the above two derivations. This rule is founded on probabilistic inference within a so-called belief network.

3 A Belief Network for Plausible Inference

Halfway through the 1980s, the belief network framework was introduced for plausible reasoning in knowledge-based systems. It provides a formalism for representing knowledge about a problem domain, or to be more precise, a formalism for representing knowledge concerning a joint probability distribution on a set of variables discerned in the said domain. In addition, the framework provides a set of algorithms for reasoning with knowledge represented in the formalism. For a general introduction to the belief network framework, the reader is referred to [Pearl, 1988]. Here, the discussion is restricted to its use in information disclosure.

3.1 A Belief Network of Index Expressions

For the purpose of plausible inference for information disclosure, a belief network of index expressions is presented. Informally speaking, a belief network comprises two parts:

- a *qualitative* part taking the form of an acyclic directed graph depicting the variables discerned in a domain as vertices and their probabilistic interdependencies as arcs, and
- a *quantitative* part taking the form of a set of (conditional) probabilities quantifying the dependencies between the variables discerned.

These two parts taken together define a unique joint probability distribution over the variables discerned; this probability distribution reflects the (in)dependencies between the variables portrayed by the qualitative part of the belief network, [Van der Gaag, 1990].

In constructing the qualitative part of a belief network of index expressions, the variables involved in the problem of information disclosure and their interdependencies have to be identified. To this end, the lithoid is taken as a point of departure. A searcher may exploit a lithoid for information disclosure by navigating over it, [Bruza, 1990]. In such a search, some of the index expressions in the lithoid are possibly relevant and some are not. Each of the index expressions of the lithoid may therefore be viewed as defining a probabilistic variable that takes one of the values *true*, that is, relevant, or *false*, that is, not relevant.

As to the dependencies between these variables, recall that the index expressions in the lithoid are partially ordered by the is-subexpression-of relation $\underline{\subset\!\!\ast}$. It follows that the probabilistic variables corresponding with these index expressions are partially ordered by this relation as well. As the lithoid captures this relation, its edges designate the dependencies between the variables. The topology of the lithoid can therefore be taken as the underlying graph of the qualitative part of the belief network. The edges of this undirected graph are assigned a direction using the inverse $\underline{\subset\!\!\ast}$-relation, expressing that belief in the relevance of an index expression is dependent upon the belief in the relevance of the index subexpressions it is built from. To conclude, we observe that the empty index expression may be omitted from the belief network as it is not information bearing. As an example, Figure 3 depicts the directed graph constructed from the lithoid shown in Figure 2.

In specifying the quantitative part of the belief network, the strengths of the dependencies between the variables discerned have to be assessed: for each variable in the qualitative part of the belief network, several (conditional) probabilities have to be provided describing the influence of values of the (immediate) predecessors of the variable on the probabilities of its own values. For specifying the required probabilities, we begin by looking at variables having no incoming arcs in the directed graph. Note that these variables correspond to single terms. For such a variable prior probabilities on its val-

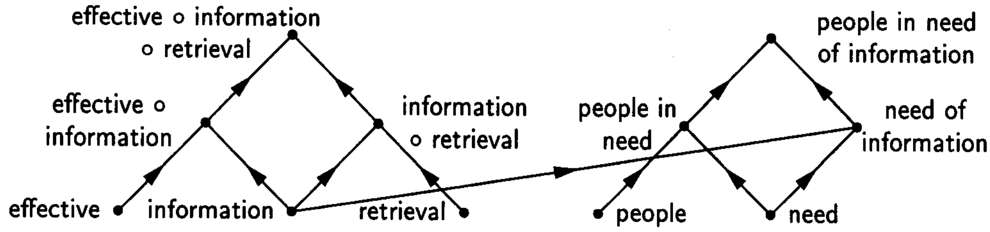

Figure 3: An Example Digraph

ues have to be specified. For the directed graph shown in Figure 3, for example, the prior probabilities

Pr(effective)
Pr(information)
Pr(retrieval)
Pr(people)
Pr(need)

have to be assessed; the complementary probabilities follow from $Pr(\neg t) = 1 - Pr(t)$, for any term t. In the context of information disclosure relative to a given set of information objects $\mathcal{O}$, it is reasonable to assume that a term that occurs frequently has a higher probability of being in a relevant object than a term that occurs infrequently. The prior probabilities on the values of a term variable may therefore be computed from the occurrence frequency of the term relative to $\mathcal{O}$. This approach to estimating the probability of relevance of a term is common in information retrieval, [Wong & Yao, 1990].

We now turn to variables in the qualitative part of the belief network having predecessors without any incoming arcs. These variables correspond to binary index expressions which are constructed from two terms via the addition of a connector. For such a variable, we have to assess the conditional probabilities of its values given values for its predecessors. One of the probabilities we have to specify is the probability of the binary index expression being relevant given that we know that its constituting terms are relevant. For the directed graph shown in Figure 3, for example, we have to assess the probabilities

Pr(effective o information | effective, information)
Pr(information o retrieval | information, retrieval)
Pr(people in need | people, need)
Pr(need of information | need, information)

Assessment of these probabilities may be based on an analysis of the frequencies of occurrence of connectors in binary index expressions. Recently such an analysis has been carried out on the titles of the Cranfield document collection, [Rosing, 1991]. This analysis revealed the connector probabilities shown in the (incomplete) table in Figure ??. Note that using these connector probabilities provides only rough estimates of the probabilities required.

Connector	*Probability*
o	0.5366
and	0.0492
as	0.0004
at	0.0348
between	0.0052
by	0.0061
for	0.0327
from	0.0039
in	0.0632
of	0.1529
on	0.0370
or	0.0026
over	0.0066
through	0.0035
to	0.0170
with	0.0248

Figure 4: Some Connector Probabilities

Beside the probabilities of binary index expressions being relevant given the relevance of their constituting terms, we have to assess three more conditional probabilities for each binary index expression: the probabilities of relevance given that one or both of its constituting terms is definitely not relevant. For example, for the binary index expression effective o information, we have to assess the additional probabilities

Pr(effective o information | ¬effective, information)
Pr(effective o information | effective, ¬information)
Pr(effective o information | ¬effective, ¬information)

These probabilities are necessarily equal to zero as a result of harbouring maximal belief in the consequences of the strict inference mechanism. For a formal proof of this property, the reader is referred to [Bruza & van der Gaag, 1992].

To conclude, attention will be focused on variables representing n-ary index expressions, $n \geq 3$. From the construction of the lithoid it will be evident that an n-ary index expression is formed by combining two index expressions of degree $n-1$ that overlap in $n-2$ terms. For such a variable, we have to assess the conditional probabilities of its values given values for its predecessors. For the probabilities of relevance of an n-ary index expression given that one or both of its constituting index subexpressions of degree $n-1$ is definitely not relevant, the property mentioned above applies once more: these probabilities are equal to zero. So, only the probability of relevance given relevance of the constituting subexpressions remains to be assessed. For the directed graph in Figure 3, for example, we have to assess the conditional probabilities

Pr(effective ∘ information ∘ retrieval |
effective ∘ information, information ∘ retrieval)

Pr(people in need of information |
people in need, need of information)

The analysis of the titles of the Cranfield document collection cited above revealed the following property: if two index expressions of degree $n-1$ having an overlap in $n-2$ terms combine into an n-ary index expression, $n \geq 3$, then this expression they combine into is unique. So, for two index expressions i and j of degree $n-1$ combining into an index expression k of degree n, the probability of k given i and j equals 1. Note that for larger document collections this probability assessment may not be accurate. However, it is expected that for larger sets of information objects there equally exists some small value of n such that for probabilistic variables representing n-ary index expressions the probability assessment mentioned above is appropriate.

The point has now been reached that for all variables discerned a set of (conditional) probabilities has been specified. These sets of probabilities and the directed graph constructed from the lithoid together constitute the *index expression belief network*. We conclude by observing that the approach presented differs from the one proposed in [Turtle & Croft, 1990] in the respect that in our approach the belief network exists purely within the realm of the object characterization language.

3.2 The Index Expression Belief Network and Plausible Inference

The basic idea now is to take the index expression belief network built from a core set of index expressions as outlined before and associate with it a rule of plausible inference for information disclosure. To this end, the set of algorithms provided by the belief network framework is exploited.

Associated with a belief network are two algorithms for making probabilistic statements: an algorithm for computing probabilities of interest from the network, and an algorithm for propagating evidence, that is, for entering a piece of evidence into the network and subsequently updating all probabilities given this evidence. These algorithms are taken to define the rule of plausible inference called *inference by probabilistic deduction*. Consider two index expressions i and j in the language $\mathcal{L}(T, C)$. Then, if $Pr(j \mid i) > 0$, j is derived from i through *inference by probabilistic deduction*, denoted as

$$i \mathrel{|\!\!\sim}_{PD} j$$

Inference by probabilistic deduction is easily combined with *modus continens*: *modus continens* is implicitly embedded in the index expression belief network. In fact, $i \vdash_{MC} j$ implies that $Pr(j \mid i) = 1$. This property follows from the topology of the graphical part of the belief network being obtained from the lithoid which in turn defines the set of index expressions derivable through *modus continens* from a core set of index expressions, and the way the probabilities in the network are defined.

4 Index Expression Belief Networks in Practice

Recently some preliminary experiments have been carried out to test the suitability of the belief network approach to plausible inference for information disclosure presented in the previous section. In performing these experiments, the IDEAL system was used: a domain-independent environment for building and reasoning with probabilistic belief networks, supporting several different algorithms for computing probabilities and for propagating evidence, [Srinivas & Breese, 1990]. We outline how the experiments were conducted.

The experiments have been performed using the Cranfield document collection. A lithoid was constructed from the first twenty-five document titles and the first three queries from this collection. The resulting lithoid comprised 1007 index expressions; 120 index expressions were shared, 114 of which comprised at most three terms. From this lithoid the qualitative part of the index expression belief network was constructed as described before. In assessing the prior probabilities for the term variables, the first five hundred titles of the Cranfield document collection were analysed; the probabilities thus acquired ranged from 0.000264 to 0.0397, for the probabilities $Pr(\mathsf{accuracy})$ and $Pr(\mathsf{flow})$, respectively. The other probabilities were assessed as outlined before.

For relevance judgement of objects, the *probability of relevance* of an object O with respect to a request q defined as

$$P_{Rel}(O,q) = max\{Pr(q \mid i) \mid i \in \chi(O)\}$$

was used. Several alternative definitions are conceivable; in fact, the experiments suggested taking relative change in belief into account. It will be a matter of further experimentation to decide on the most appropriate scheme for relevance judgement.

The basic idea of applying *inference by probabilistic deduction* is to enter an index expression from an object characterization as evidence into the belief network and propagate its impact on all other variables; thereafter, the (updated) probability of the request is computed from the network. This approach requires for a single request performing evidence propagation all over again for each index expression appearing in an object description. As evidence propagation is costly from a computational point of view, another approach was used in the experiments requiring only one evidence propagation for a single request: instead of entering an index expression from an object description as evidence and examining the probability of relevance of the request, the request was entered as evidence and the probabilities of the index expressions were inspected. Note that for a request q and an index expression i this approach delivers the probability $Pr(i \mid q)$ instead of the desired probability $Pr(q \mid i)$. However, $Pr(q \mid i)$ may be computed from $Pr(i \mid q)$ using Bayes' Rule:

$$Pr(q \mid i) = \frac{Pr(i \mid q) \cdot Pr(q)}{Pr(i)}$$

provided that the prior probabilities $Pr(i)$ and $Pr(q)$ of the index expression i and the request q, respectively, are known. In the experiments, these prior probabilities were computed from the belief network whenever needed. For practical application, however, it will be more efficient to compute all prior probabilities beforehand and store them with the variables in the network.

From the experiments, it was concluded that the plausible inference mechanism based on the index expression belief network seems to foster precision. The results obtained from the preliminary experiments are discussed in more detail in [Bruza, 1993].

A big question mark left by the experiments was whether the belief network approach to plausible inference as discussed above can be scaled to real-life information disclosure applications. Even with the restricted belief network (the result from the first optimization discussed below), propagation of a request took approximately forty minutes in the IDEAL system. The stark reality is that real-life information disclosure applications would be based on index expression belief networks containing hundreds of thousands, if not millions, of variables. It was concluded that for belief networks of this size straightforward use of the existing algorithms will not yield an efficient enough plausible inference mechanism for realistic information disclosure: if belief networks are to be employed, very efficient, special-purpose algorithms need be devised. The groundwork for such algorithms will be laid in the following section.

5 Lessening the Computational Expense

Fortunately, an index expression belief network exhibits some special properties that can be exploited to lessen the computational expense involved in plausible inference. These properties arise from its embedding the strict inference mechanism.

Two optimizations of the basic inference mechanism are proposed. The first optimization is to restrict the size of the index expression belief network. As described in Section 3, the qualitative part of the belief network is constructed from the lithoid of index expressions; the probabilities constituting its quantitative part are assessed from an analysis of the titles of the Cranfield document collection. Now recall that this analysis revealed that if two index expressions i and j of degree $n-1$ combined into an n-ary index expression k, $n \geq 3$, then they did so uniquely, leading to the probability assessment $Pr(k \mid i, j) = 1$ for the corresponding variables. Since the probability assessments for all these variables are the same, there is no need to represent them separately and explicitly in the network. This observation suggests restricting the belief network to variables representing small index expressions only: in such a restricted network evidence propagation is much faster as it involves less variables. A simple scheme is then used for computing probabilities of relevance for larger index expressions exploiting the special properties of their accompanying probability assessments.

The effect of restricting the size of the index expression belief network can be envisioned by the lithoid as follows. A lithoid resembles a mountain range, each peak corresponding to a power index expression. Restricting the lithoid to index expressions of, for example, three terms or less implements lopping off the peaks of this mountain range leaving only a common base. This common base in fact captures the part of the lithoid that is most interesting for plausible inference as it contains roughly 95% of shared index expressions, [Bruza, 1993]; shared index expressions allow spreading of evidence throughout the associated belief network. This optimization was applied to the lithoid of 1007 index expressions constructed in our experiments; the restricted lithoid comprised 393 index expressions, only.

The second optimization proposed is to limit evidence propagation to a relevant subnetwork of the index expression belief network. This optimization was motivated by the observation that index expression belief networks are much larger in width than they are in height. This feature suggests that the number of variables whose probabilities are affected by evidence propagation is small compared to the number of variables whose probabilities are not affected; in fact, in an experiment with the Cranfield document collection the ratio of numbers of affected and unaffected variables equalled 0.05. Because it will suffice to update the probabilities of the affected variables only, much computational effort in evidence propagation can be saved provided that these variables can be identified efficiently.

Consider a request q. The set of variables in the index expression belief network whose probabilities are affected by entering q as evidence is defined as

$$victims(q) = \{i \mid Pr(i \mid q) \neq Pr(i)\}$$

The set of unaffected variables is defined as

$$immunes(q) = \{i \mid Pr(i \mid q) = Pr(i)\}$$

Note that for each variable i in the belief network, we have either $i \in victims(q)$ or $i \in immunes(q)$. The aim now is to identify the set $victims(q)$ from the qualitative part of the index expression belief network only, that is, without any probabilistic computation.

It will be evident that by entering the request q as evidence into the belief network, its own probability of relevance will be set to 1. From the strict inference mechanism being embedded in the belief network it follows that for all index subexpressions of q the probability of relevance will be updated to unity. So, for each variable corresponding to an index expression $i \in \mathcal{P}(q) \setminus \{\epsilon\}$, we have that $Pr(i \mid q) = 1$; in the sequel, we will consider each such index expression as a piece of evidence. It is concluded that $\mathcal{P}(q) \setminus \{\epsilon\} \subseteq victims(q)$. In addition, all variables corresponding to index expressions that can be obtained through refinement from an index expression $i \in \mathcal{P}(q) \setminus \{\epsilon\}$ belong to the set $victims(q)$. It can be shown that all other variables are independent of the request q and are therefore included in the set $immunes(q)$. To conclude, the set $victims(q)$ equals

$$victims(q) = \{j \mid i \mathrel{|\!\sim}_{PR} j,\ i \in \mathcal{P}(q) \setminus \{\epsilon\}\ \}$$

This set can easily be computed using the is-subexpression-of relation $\underline{\circledast}$. The basic idea of the set $victims(q)$ is schematically represented in Figure 5.

For updating the probabilities of the values of the variables in the set $victims(q)$, a standard algorithm for evidence propagation can be used. However, the probability assessments associated with the variables in the index expression belief network and the interdependencies between

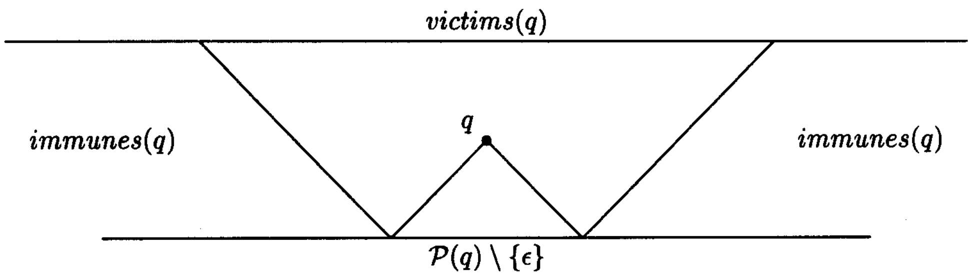

Figure 5: A Subnetwork for Limited Evidence Propagation

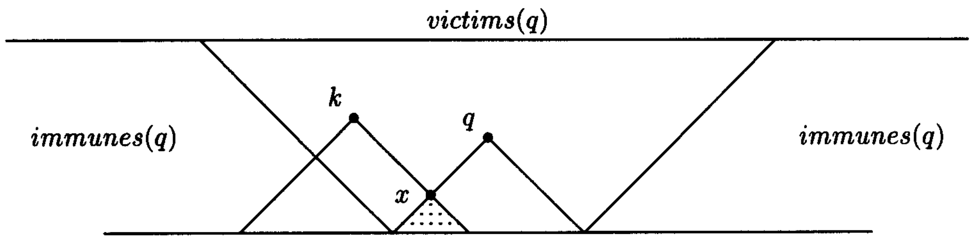

Figure 6: Efficient Evidence Propagation

these variables once more allow for more efficient computation of probabilities of relevance.

Referring to Figure 6, consider a variable $k \in victims(q)$. The only evidence that pertains to k directly, originates from the vertices in the set $\mathcal{P}(x)\backslash\{\epsilon\}$; all other vertices in $\mathcal{P}(q)\backslash\{\epsilon\}$ have no direct influence on k because k is independent of them. By exploiting the fact that the strict inference mechanism is embodied in the belief network, it can be proven that

$$Pr(k|q) = \frac{Pr(k)}{Pr(x)}$$

Using this result means that there is a single computation involved for updating the probability of each victim k of q. So this optimization yields a time complexity of plausible inference that is linear in the number of victims of a given request. A paper presenting full details on this optimization is in preparation.

In conclusion, we observe that the optimizations proposed exploit the features of the strict inference mechanism to a large extent and therefore cannot be applied to more general belief networks; in fact, evidence propagation can have far reaching effects in general.

6 Conclusions

In this paper, a context-sensitive plausible inference mechanism for information disclosure has been proposed, based on the index expression belief network. The results obtained from the preliminary experiments with this inference mechanism are encouraging. These experiments, however, are too small-scaled to provide real insight into the potential effectiveness of our approach; we hope to report on larger experiments in the near future. The feasibility of the belief network approach for driving plausible inference in information disclosure remains an area of continuing interest.

References

[Bruza, 1990] P.D. Bruza (1990). Hyperindices: A novel aid for searching in hypermedia *Proceedings of the European Conference on Hypertext - ECHT 90*, pp. 109 - 122.

[Bruza, 1993] P.D. Bruza (1993). *Stratified Information Disclosure: A Synthesis between Information Retrieval and Hypermedia.* PhD thesis, University of Nijmegen.