

Risk Information Extraction and Aggregation

Experimenting on Medline Abstracts

Léa Deleris¹, Stéphane Deparis¹, Bogdan Sacaleanu¹, and Lamia Tounsi²

¹ IBM Research – Ireland, Dublin

{lea.deleris, stephane.deparis, bogdan.sacaleanu}@ie.ibm.com

² Dublin City University

lamia.tounsi@computing.dcu.ie

Abstract. By exploiting advances in natural language processing, we believe that information contained in unstructured texts can be leveraged to facilitate risk modeling and decision support in healthcare. In this paper, we present our initial investigations into dependence relation extraction and aggregation into a Bayesian Belief Network structure. Our results are based on a corpus composed of MEDLINE® abstracts dealing with breast cancer risk factors.

Keywords: Probability, influence, Natural Language Processing, risk analysis, information extraction, expert aggregation, Bayesian networks.

1 Introduction

Bayesian belief networks (BBNs) [1] have been applied within a variety of contexts, including engineering, computer science, medicine and bioinformatics. They are a popular tool for risk analysis modeling and decision support systems [2, 3]. While BBNs can be constructed automatically from structured data, the amount of human effort required to construct a BBN manually, either from expert opinion or based on a literature review, may be impractical on a large scale.

However, there is a wealth and growing amount of information in unstructured format that could be leveraged to facilitate the creation of BBNs (Medical academic knowledge being one example). The challenges associated with unstructured texts are many. Relevant information is (i) sparse, (ii) scattered among a large amount of irrelevant sentences and (iii) ambiguous due to the richness of the human language. Beyond the extraction of that information, synthesizing it in a coherent model presents its own challenges, akin to those associated with the construction of risk models from multiple experts, in particular diverging opinions and partial information.

Our goal is nevertheless to investigate how current advances in natural language processing techniques could be used to address the scalability problem of extracting relevant pieces of information to build BBN-based risk models from unstructured texts. Note that our objective is not to replace experts but rather to facilitate their task. Therefore the focus is on extracting the information and creating initial versions of BBNs which are then to be edited by domain experts.

In this paper we describe our overarching approach to construct BBN from text followed by a detailed description of extraction and aggregation of structural information. The extraction and aggregation of the quantitative information is not addressed here.

2 Description of the Research Problem

2.1 Overarching Approach

This section presents the general approach that we have defined to construct BBNs from unstructured texts. BBNs are composed of a graph, whose nodes (or vertices) represent random variables and whose arcs (or edges) capture dependence statements (independence, the stronger statements, is derived from the absence of arcs). Each node is associated with a conditional probability table which provides numerical information about the strength of the dependence. People commonly refer to the graph (nodes and arcs) as the graphical layer or structure of the BBN while the term quantitative layer pertains to the conditional probability tables (CPTs) of the nodes. The construction of BBN *based on expert knowledge* typically follows three main steps:

1. Identification of variables (nodes), possibly definition of the states (although this does not need to be done till before the third step)
2. Definition of the network structure, i.e., the dependence and independence relations among the variables (arcs)
3. Specification of the parameters of the CPTs.

How does this process translate into text-based extraction of expertise? Because the elements of the quantitative layer depend on the graphical layer (the structures of the conditional probability tables depend on the parents of each node) it is reasonable when obtaining information from experts to determine the structure of the BBN before populating it with quantitative information. This constraint does not apply when information is extracted from a set of unstructured texts because the extraction step is not tailored to a given structure. Indeed it is unlikely that the probability statements that can be extracted will perfectly match the required CPT inputs. Limiting the search for quantitative information solely to the CPTs would result in discarding relevant information. We assume therefore that the dependence relation extraction and the probability statement extraction can be done in parallel as independent tasks.

By contrast, the output of the transformation of the extracted information into a BBN can vary depending on the sequence of subtasks. Rather than constructing a BBN for each source and then combine those together, we chose to first aggregate all dependence information into a network structure and thereafter to populate the associated CPTs based on all the probability statements extracted. We think that pooling information together as early as possible will reduce aggregation challenges: When probability and dependence statements are few at each source level but aggregated over many sources, the individual BBNs for each source are likely to be dissimilar.

Our suggested methodology is summarized on Fig. 1. We start by building the information base through the extraction of (i) structural information in the form of dependence and independence relations among risk factors (e.g. Age at Menarche \rightarrow Breast Cancer Incidence) and (ii) quantitative information in the form of probability statements (e.g., $P(\text{Lifetime Breast Cancer} \mid \text{Malaysian Women}) = 0.05$). The content of the information base is then processed to generate variables, states, aggregated network structure and finally parameters for the CPTs.

In this paper we describe in more details the dependence extraction and the construction of the aggregated network structure. Despite not having achieved full automation of the extraction process, we are able to shed some light on the feasibility of both extraction and aggregation and on some of the related challenges. Regarding variable and state identification, we intend to address them in future research and resort to simple heuristics in the meantime. Specifically, different terms are considered as different variables and variables are considered as binary so that state identification is not required. In the future, we will leverage knowledge bases such as UMLS along with machine learning methods to cluster terms together into variables. Regarding quantitative information, we have defined and evaluated algorithms for extraction and evaluation but chose not to present them due to space constraints.

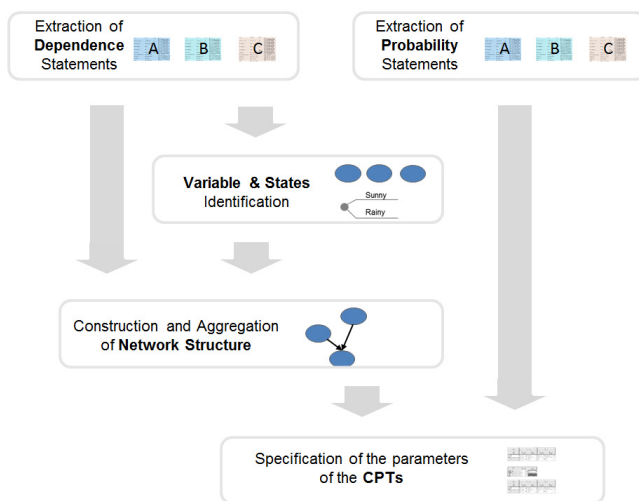


Fig. 1. Steps involved in building BBN from Text

2.2 Related Research

To the best of our knowledge, there is limited research that focuses on building BBNs from unstructured text. There has been some early work related to building BBN from structured text information. [4] presents a graph grammar which takes as input structured information (disease, treatment, test) and constructs an influence diagram (BBN augmented by decision nodes) from it. Similarly, [5] leverages the information

contained in Medical Subject Headings (MESH) and subheadings in MEDLINE citations to build the structure of an influence diagram. Both approaches are specific to the medical domain as they rely on pre-determined keywords to guide the structuring of the network. [6] appears to be the first paper to specifically focus on presenting an NLP framework for building BBN from unstructured texts. The authors, coming from the natural language processing community, restrict their search to causal relation extraction to determine the edges of the network. While they leverage semantic information extracted from WordNet for variable identification, they provide a limited discussion on the construction of the graph, ignoring challenges related to aggregation of all the causal information into a coherent BBN such as the creation of cycles and the need to distinguish between direct and indirect relations. Our approach, while similar in spirit, builds more significantly upon the decision science literature. In that domain, there has been a significant amount of work in combining expert opinions when those opinions are probabilities, for instance [7], yet the combination of Bayesian network structures provided by experts is seldom considered. Notable exceptions are [8-10] which provide practical and theoretical perspectives on the fusion of Bayesian networks. One important requirement of these research efforts is that input are composed of true Bayesian networks, meaning directed dependence relations among random variables, while our input would be better described as degenerate versions of Bayesian network in the form of undirected dependence and independence statements among pairs of random variables. The works of [11-12], which we will discuss in section 4, are better aligned with our task as relying on looser input requirements.

2.3 Focus on Healthcare

So far, we have concentrated our efforts in applying the above framework in the healthcare domain. Indeed, BBNs have long been advocated as a useful decision and risk modeling framework in medicine [13] and used in several decision support systems [14-16] yet the difficulty of building BBNs from scratch has limited widespread adoption [17]. In fact, [17] reports that the process of building the structure of model for therapy selection for the treatment of the cancer of the esophagus required 11 sessions with 2 experts and one knowledge engineer, each session lasting 2-4 hours and requiring about 20 hours of preparation. Our focus on automating the extraction and aggregation of relevant information seeks to address this practical challenge.

This paper focuses specifically on breast cancer, simply motivated by the large amount of academic papers published on the topic each year and the fact that it is a disease that is well understood. We created a *reference information base* (called gold standard in the natural language processing community) from about 300 MEDLINE® abstracts selected according to the query: “KW¹:breast cancer AND parity” over the past 5 years. This reference information base consists of the manually generated output of the two extraction tasks. It is used to understand the characteristics of the elements to be extracted, to develop machine learning algorithms for extraction and, in this paper, as proxy for automatically extracted information (for aggregation).

¹ KW stands for Keyword(s).

3 Structural Information Extraction

3.1 Defining Dependence and Independence

Structural information extraction involves identifying in texts information indicating dependence or independence among variables. Dependence and independence here are to be understood as defined by probability theory where A depends on B iff $P(A) \neq P(A|B)$. As independence statements are seldom, we focus predominantly on dependence in the remaining of this paper and define independence as the negation of dependence. Future research will focus on understanding the specificities of independence statements to increase the chances of extracting such statements.

Our objective in the extraction step is to transform a sentence into set of dependence relation entries structured as follows:

- One Variable A and One Variable B (compulsory)
- Influence terms (compulsory)
- A modifier (optional)
- A context variable (Optional)
- A negation (Optional)

While dependence and independence of variables are symmetric relations, we make a loose distinction between Variable A and Variable B, where Variable A captures the causing factor and Variable B the influenced factor. We make influence terms compulsory because from a language perspective, they serve as the cornerstone of the dependence relation. Examples of influence terms include: “associated with”, “reduction”, “correlated”, “higher” or “likely”. The modifier can provide nuances about the strength of the statement (e.g., “significantly”, “may”, and “positively”), the context variable limits the population to which the statement applies, and the negation element enables to capture independence. Other optional variables could be added, typically meta-data such as authorship, venue, publication year and original language. As an example, from the sentence “For endometrial cancer, body mass index represents a major modifiable risk factor; about half of all cases in postmenopausal women are attributable to overweight or obesity.”, we extract two structured relations:

- Variable A: body mass index – Variable B: endometrial cancer – Influence Term: risk factor – Modifier: major
- Variable A: overweight or obesity – Variable B: cases – Influence Term: attributable – Context: postmenopausal women.

Our first step is to characterize the elements to be extracted (so far variables and influence terms) and then to develop machine learning approaches using features derived from the characteristics uncovered. Dependence information extraction has many similarities with relation extraction from clinical texts whose focus has been on protein-protein interactions or protein-gene interactions, with less attention being paid to the extraction of other types of relationship [18-20]. Our goal here is different as we approach a new type of relation (risk) and do not have any predefined restrictions for the variables beside the medical domain.

3.2 Characterizing Elements

Using 30 abstracts² from our information base, we analyzed the syntactic and semantic characteristics of variables. Syntactically, variables are noun phrases (NP) with different levels of complexity: Variables A are mostly specific displaying context information (e.g. *age at menarche*, *duration of breast feeding*, *personal history of breast cancer*), while Variables B are rather concise (e.g. *breast carcinoma*, *mammographic density*). Semantically, using UMLS Metathesaurus semantic types³ referred to in biomedical texts, we observed the following patterns:

- Variables A: [Finding] / [Therapeutic or Preventive Procedure] / [Organism Function] / [Disease or Syndrome] / [Gene or Genome] / [Organism Attribute]
- Variables B: [Neoplastic Process] / [Qualitative Concept] / [Disease or Syndrome]

Regarding influence terms, they appear to belong to a limited controlled language. From our full information base, we identified 1000 relations representing only 117 unique terms out of a vocabulary of about 10000 words. We report in Table 1 frequency counts of the 10 most frequent roots of those influence terms along with ratio of occurrences in a dependence relation by the total occurrences. Note that the five most common roots account for about 50% of the influence terms occurrences and the 10 most frequent for almost two third. Note also that for many roots, their occurrence is highly indicative of the presence of a dependence relation (fourth column).

Table 1. Frequency Counts of Most Frequent Influence Terms

Root	Occurrences in relations	Frequency among identified influence terms	Occurrences in relation / Total occurrences
associate	333	34.5%	64.8%
increase	76	7.9%	37.4%
risk	62	6.4%	9.2%
relate	42	4.3%	36.5%
reduce	33	3.4%	55.0%
influence	26	2.7%	54.2%
likely	23	2.4%	71.9%
effect	22	2.3%	27.2%
correlate	20	2.1%	62.5%
high	17	1.8%	22.7%

² Those corresponds to the first 30 abstracts that were returned from Medline with our query.

³ UMLS Metathesaurus semantic types enable to categorize the concepts listed in the Metathesaurus. There are 133 semantic types which are grouped around the following broad categories: organism, anatomical structure, biologic function, chemical, physical object, idea or concept. See [22] for a more in-depth description.

3.3 Extraction Algorithm

Consequently, using our information base we will build a dictionary of influence terms. The main challenge that we currently face is to determine for each occurrence of such terms whether it does correspond to a dependence relation and to extract the associated variables. As commonly done in natural language processing, we rely on machine learning for this task. Each term in a text will represent a candidate to be classified using features ranging from lexical ones (e.g., distance to influence term), to syntactic ones (e.g., part-of-speech, grammatical function, constituent type), to semantic ones (e.g., UMLS type). We plan to experiment with different classification algorithms for each variable type (A and B). Since the training set is small, generative classifiers like Naive Bayes are preferred initially as they are less subject to overfitting. Afterward, we plan to use additional unlabeled data with the Naive Bayes classifier and combine them in a semi-supervised learning approach. The resulting extended set of labeled data will be considered for learning with ensemble methods for decision trees that tend to outperform other algorithms for classification problems.

However, there are several steps that have to be undertaken before using the current information base for learning. Medical abstracts have to be preprocessed to deal with two linguistic challenges highly present in this domain, namely anaphoric reference and ellipsis. Referencing happens when the authors refer back to previously mentioned entities to avoid repetition (e.g., it referencing BRCA⁴), while ellipsis occurs when, after a more specific mention, words are partially or completely omitted when the phrase needs to be repeated.

Similarly we need to post-process the variables extracted to handle terms referring to an underdefined set of variables (e.g., physician characteristics, personal factors) and to cluster those referring to the same risk factor (e.g., smokers, smoking, cigarettes, tobacco consumption) into a single variable. This corresponds to our variable identification step which we will undertake using general lexical resources like the Roget Metathesaurus or specialised ones as the UMLS Metathesaurus and borrowing from [6]. The final output of the extraction and variable identification steps would be a set of cleaned evidence statements representing dependence and independence between variables rather than terms. Those are the input for the information aggregation described hereafter.

4 Structural Information Aggregation

4.1 General Problem Description

One way to aggregate a cleaned set of evidence statements into a BBN structure is:

1. to reconcile statements involving the same pair of variables A and B. The simplest approach being a majority vote, which can be modulated by adding weights to each statement, based on modifiers, publication year of the paper (higher weights for more recent papers) or any other relevant information.

⁴ BRCA is a mutation of the genes BRCA1 and BRCA2 associated with higher risk of developing breast cancer.

2. to aggregate reconciled statements into a BBN structure, i.e., into a directed acyclic graph. One process to do so is to first create the undirected graph associated with the reconciled statements, assuming lack of information is understood for instance as independence (but the reverse convention is as valid, if not more) and second to orient the edges by assigning a random order to the variables thus ensuring that no cycle is created in the process.

The specific methods described in each step above are only the most straightforward solutions provided for the sake of illustration. More sophisticated approaches on the majority vote of the adjacency matrix of the graph, adapted for instance from [11] could also be applied. In addition, a Bayesian approach as described by Richardson and Domingos [12] is a valid alternative where, instead of having as input BBN structures from multiple experts, we have evidence statements from multiple papers. Finally yet another approach can be used based on an adaptation of the PC Algorithm [21]. While the PC Algorithm is based on independence tests, it can be adapted by replacing those tests with the independence statements extracted from the evidence statement extraction after reconciliation for conflicting opinions.

4.2 Our Method

We adapt Richardson and Domingos (RD hereafter) method for aggregation of multiple BBNs provided by human experts (and data, but that is not relevant in this paper). In their approach, each expert gives a description of the dependencies among a fixed set of variables, which is used to compute the most probable structure doing Bayesian updating. Some assumptions are made to simplify the computations, in particular that experts statements are independent from one another both across experts and across pairs of variables. As will become clear in the following paragraph, we make similar independence assumptions. However, the main difference in our method is that we assume that the information provided by sources (abstracts for us, experts for RD) is *not* directed. This simplifies the specification of the likelihood function yet implies that the output of the model is non-directed. We thus need to orient the edges to obtain a full BBN.

We now describe our approach, starting with some notation:

- Let $X = \{x_1, \dots, x_N\}$ denote the set of variables in the information base, arbitrarily ordered.
- Let G denote a non-directed graph over A , we choose to represent $G = (g_1, \dots, g_j, \dots, g_{N'})$ as a vector of size $N' = N \cdot (N - 1)/2$ where g_j corresponds to the j th variable pair for some arbitrary ordering of all possible pairs. $g_j = 1$ if there is an edge between the variables forming the j th pair and $g_j = 0$ otherwise.
- Let M denote the number of sources (texts) in our information base and $E_m = (e_{m1}, \dots, e_{mj}, \dots, e_{mN'})$ represent the evidence statements from the m th source where $e_{mj} = \emptyset$ means that no statement has been made about the j th pair, $e_{mj} = D$, means a dependence statement has been made and $e_{mj} = I$ means an independence statement has been made. Finally we denote by $E = (E_1, \dots, E_m, \dots, E_M)$ the full content of the information base.

Our objective is to evaluate the posterior distribution of the underlying non-directed graph $P(G|E)$, which by Bayes theorem is $P(G|E) = \alpha P(G)P(E|G)$. We need to specify the prior $P(G)$ and the likelihood function $P(E|G)$. For simplicity, we assume that the presence of an arc between a pair of variables is independent from the other pairs, thus $P(G) = \prod_{j=1}^{N'} P(g_j)$. In addition, we assume that sources are conditionally independent given graph G so that $P(E|G) = P(E_1, \dots, E_m, \dots, E_M | G) = \prod_{m=1}^M P(E_m | G)$. Finally, we assume that evidence statement from source m about pair j is conditionally independent from the other evidence statements from source m and depends only on g_j and not $\{g_l\}_{l \neq j}$. This leads to

$$P(G|E) = \alpha P(G)P(E|G) = \alpha \prod_{m=1}^M \left(\prod_{j=1}^{N'} P(g_j) \cdot P(e_{mj} | g_j) \right) \quad (1)$$

We acknowledge that we are making strong independence assumptions. We will revisit those assumptions in future research yet the current model has value in the sense that it provides a benchmark for more sophisticated models. Moreover, in several other situations, it has been found that naïve models such as ours provide a useful first approximation (e.g., naïve Bayes classification, Markov chains modeling of complex systems, aggregation of expert opinions about probabilities).

Applying (1) enables us to derive the most likely non directed structure G^* associated with the information base E . To orient the arcs, we rely on a user-defined order of the variables, which we obtained by clustering the diseases (and associated UMLS concepts) at the end of the list so that arcs would typically go from factors/symptoms to disease. This is consistent with our findings in section 3.2 related to Variables A and B.

4.3 Experimental Results

Data. From the reference information base of about 1000 relations, we selected 198 cleaned evidence statements, limiting ourselves to simple concepts that were unambiguous to non-medically trained researchers. “*Her2-overexpressing cases*”, “*basal-like molecular subtype*” were thus discarded, as were references to subtypes of breast cancer (Triple-negative, ER+/GR+). To reduce the number of variables to a manageable size, we selected the 18 most frequent which led us to keep 96 evidence statements. As expected given the limited size of our text data, we only have very partial information about the relationships among those variables. In our example, out of the 153 possible pairs, we obtained

- 18 pairs associated with only one evidence statement (11 pairs with one dependence statement and 7 pairs with one independence statement),

- 10 pairs associated with multiple concurring dependence statements (none for independence statements),
- 5 pairs associated with conflicting evidence statements (interactions between breast cancer and respectively age at menarche, breastfeeding, hormone replacement therapy, fertility, and single-nucleotide polymorphism).

We believe however that expanding the corpus may not resolve the lack of full coverage. Medical research focuses mostly on interactions between factors and diseases rather than among factors.

Parameters. In our experiment, we assume that each pair of variables independently presents a probability p_0 of being connected and set p_0 to 0.05, similarly to RD. We similarly assume that all likelihood parameters $P(e_{jm}|g_j)$ are identically distributed according to Table 2. Through this parameterization, lack of mention of a relation (\emptyset) is neutral in the sense that observation of “no mention” leaves the prior probability unchanged. While it has no effect on the computations, for the sake of completeness we set $p_\emptyset = 0.995$ (based on observed value in our information base). We set $p_D = p_I = 0.9$, a somewhat arbitrary estimation of the reliability of medical research on which we perform sensitivity analysis.

Table 2. Likelihood Model

		e_{mj}		
		\emptyset	D	I
g_j	1	p_\emptyset	$(1 - p_\emptyset)p_D$	$(1 - p_\emptyset)(1 - p_D)$
	0	p_\emptyset	$(1 - p_\emptyset)(1 - p_I)$	$(1 - p_\emptyset)p_I$

Output. Applying the method described above, the BBN structure that we would obtain is presented on Fig. 2a. where labels indicate the associated posterior probabilities that there is an arc between the pair of variables j i.e. $P(g_j = 1|E)$. In fact, as our goal is not to provide a definite model but rather an easily understandable synthesis of the extracted information that tends toward a BBN representation, we deviate from providing the most likely structure. Instead, we report two graphs derived from the posterior probabilities as follows: If $P(g_j|E) > \beta P(g_j)$ then we show an arc on Fig. 2a, if $P(g_j|E) < P(g_j)/\beta$ then we show an arc on Fig. 2b (with $\beta = 3$ in this illustration). Any pair of variables that is not connected in either graphs is therefore in the unknown category (which could be represented graphically). In this perspective, the order of the variable used for arc orientation is not critical; its role is mostly to avoid cycles and to generate an intuitive output.

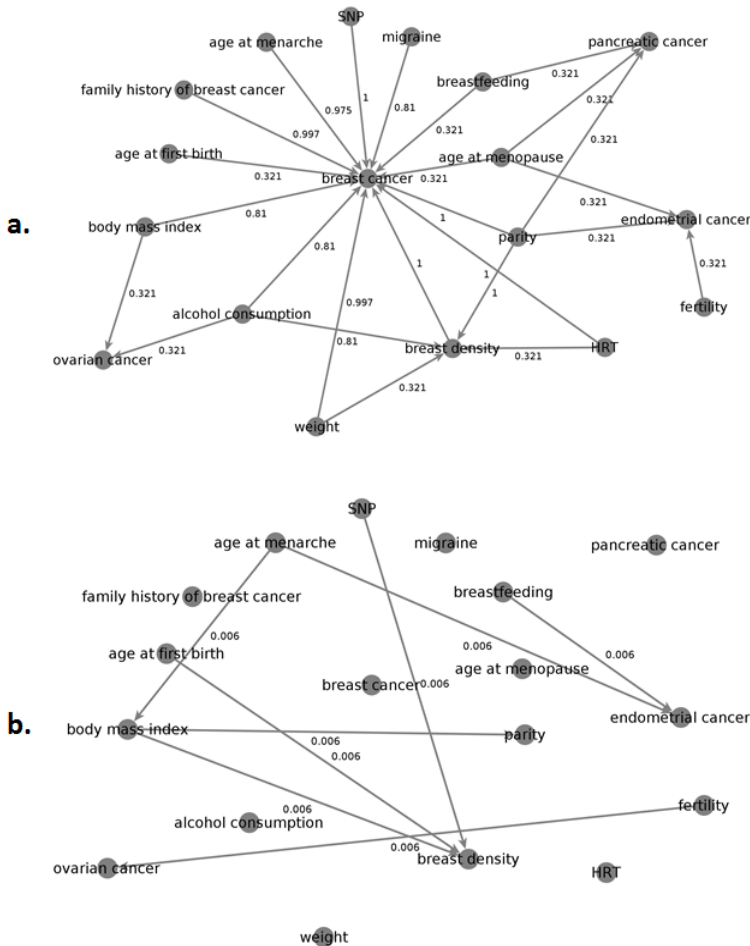


Fig. 2. Aggregation Output: Dependence Relations (top) and Independence Relations (bottom)

Overall, the information presented on Fig. 2 is designed to be challenged by a human expert. Variable order could be modified; logical relations (such as an arc between fertility and parity) could be added. In a real implementation, the expert would also be able to select an arc and view the underlying information for validation. In addition to the graphs, pairs of variables having strong support (multiple concurring relations) or high conflict (multiple conflicting relations) would be highlighted. Furthermore, we envision querying the user for validation of non-consensual statements and indirect influences (i.e. when we have both $A \rightarrow B \rightarrow C$ and $A \rightarrow C$), asking whether the second influence really exists or was generated from a simplification of the indirect one. For the BBN presented in Fig. 2 for instance, we would question whether the (protective) effect of parity on breast cancer is limited to breastfeeding.

One question that can be raised is about the benefits of the sophisticated Bayesian approach which seems to behave very much like majority vote. This is in part due to

our strong independence assumptions. In future research, we will investigate how social networks about authorship (co-authors / referred and referring authors) could be used to build a simple dependence model about text sources so as to make the benefits of the Bayesian updating method more salient.

A Note on Sensitivity Analysis. We performed sensitivity analysis on the parameters, using graph edit distance to measure the effect of parameters changes. If p_0 stays within [0.01-0.2], then the output graph remains exactly the same, thus displaying reasonable robustness. In addition, the values of p_D and p_I only influence the arcs supported by multiple *conflicting* statements, which would nonetheless be singled out for appraisal by a human expert. More care, however, needs to be spent in defining the default values of those parameters.

5 Conclusion

This paper outlines our general framework for facilitating the construction of Bayesian networks from unstructured texts, focusing initially on the medical domain. Our objective is to reduce the need for human intervention in just the same way that machine translation of a text provides a human translator with an imperfect but labor-saving first draft. We discuss more specifically the problem of extracting and aggregating dependence information for which we propose, if not fully defined algorithms, a detailed description of the steps to follow. The main shortcomings of our current analyses are (i) the lack of final algorithm for extraction and (ii) the limited size of our reference set for aggregation. Therefore we cannot reach definite conclusions. In particular, we can expect that the input to aggregation will be several orders of magnitude larger once the extraction step is fully automated. We feel nonetheless that the findings discussed in this paper are encouraging in that they confirm the feasibility of our endeavor and illustrate its usefulness.

Acknowledgements. The authors gratefully acknowledge the financial support provided by the Irish Industrial Development Agency under reference 199954.

References

1. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kauffman, San Francisco (1988)
2. Pietzsch, J.B., Paté-Cornell, E.: Early technology assessment of new medical devices. *International Journal of Technology Assessment in Health Care* 24, 37–45 (2008)
3. Deleris, L.A., Yeo, G., Siever, A., Paté-Cornell, E.: Engineering risk analysis of a hospital oxygen supply system. *Medical Decision Making* 26, 162–172 (2006)
4. Egar, J.W., Musen, M.A.: Automated modeling of medical decisions. In: *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pp. 424–428 (1993)

5. Zhu, A., Li, J., Leong, T.: Automated knowledge extraction for decision model construction: A data mining approach. In: *AMIA 2003 Symposium Proceedings*, pp. 758–762 (2003)
6. Sanchez-Graillet, O., Poesio, M.: Acquiring Bayesian networks from text. In: *Proceedings of LREC (2004)*
7. Clemen, R.T., Winkler, R.L.: Aggregating probability distributions. In: Edwards, W., Miles, R., von Winterfeldt, D. (eds.) *Advances in Decision Analysis*, pp. 154–176. Cambridge University Press, Cambridge (2007)
8. Matzkevitch, I., Abramson, B.: The topological fusion of Bayes Nets. In: Dubois, D., Wellman, M.P., Ambrosio, B., Smets, P. (eds.) *Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, San Francisco (1992)
9. Sagrado, J., Moral, S.: Qualitative combination of Bayesian networks. *International Journal of Intelligent Systems* 18, 237–249 (2003)
10. Zhang, Y., Yue, K., Yue, M., Liu, W.: An approach for fusing Bayesian networks. *Journal of Information and Computational Science* 8, 194–201 (2011)
11. Rush, R., Wallace, W.A.: Elicitation of knowledge from multiple experts using network inference. *IEEE Transactions on Knowledge and Data Engineering* 9, 688–696 (1997)
12. Richardson, M., Domingos, P.: Learning with knowledge from multiple experts. In: Fawcett, T., Mishra, N. (eds.) *Proceedings of the Twentieth International Conference on Machine Learning*. Morgan Kaufmann, Washington (2003)
13. Pauker, S.G., Wong, J.B.: The influence of influence diagrams in medicine. *Decision Analysis* 2, 238–244 (2005)
14. Warner, H.R., Haug, P., Bouhaddou, O., Lincoln, M., Sorenson, D., Williamson, J.W., Fan, C.: Iliad as an expert consultant to teach differential diagnosis. In: *Proceedings of the Annual Symposium on Computer Applications in Medical Care*, vol. 154, pp. 371–376 (1988)
15. Hoffer, E., Feldman, M., Kim, R., Famiglietti, K., Barnett, G.: Dxpplain: patterns of use of a mature expert system. In: *AMIA 2005 Symposium Proceedings*, pp. 321–325 (2005)
16. Fuller, G.: Simulconsult. *Journal of Neurology, Neurosurgery and Psychiatry* 76, 10 (2005), <http://www.simulconsult.com>
17. van der Gaag, L.C., Renooij, S., Witteman, C.L.M., Aleman, B., Taal, B.F.: How to elicit many probabilities. In: Laskey, K.B., Prade, H. (eds.) *Proceedings of the Fifteenth Conference on Uncertainty*, pp. 647–654. Morgan Kaufmann Publishers, San Francisco (1999)
18. Chowdhary, R., Zhang, J., Liu, J.: Bayesian inference of protein-protein interactions from biological literature. *Bioinformatics* 25, 1536–1542 (2009)
19. Bui, Q.C., Katrenko, S., Sloot, P.M.A.: A hybrid approach to extract protein-protein interactions. *Bioinformatics* 27, 259–265 (2011)
20. Miwa, M., Sætre, R., Kim, J.-D., Tsujii, J.: Event extraction with complex event classification using rich features. *Journal of Bioinformatics and Computational Biology* 8, 131–146 (2010)
21. Spirtes, P., Glymour, C., Scheines, R.: *Causation, prediction, and search*. Lecture Notes in Statistics. Springer (1993)
22. UMLS® Reference Manual [Internet]. Bethesda (MD): National Library of Medicine (US); Semantic Network, <http://www.ncbi.nlm.nih.gov/books/NBK9679/>