Abstract

We examine some issues relating to the hypothesization of relevant models of a situation when the domain knowledge is primarily probabilistic in nature. Our formalism seeks consistent, minimal, sufficiently large, and relevant situation-descriptions when a preference criterion is specified.

1 Introduction

We outline a formalism for constructing models of hypothetical situations that can possibly arise in a domain. The formalism is for those domains in which the domain knowledge is primarily probabilistic in nature. For a very small example, we consider a domain containing attribute X with possible values (\(a, b, c\)) and attribute Y with possible values (\(l, m, n\)). A particular value-assignment to each attribute forms a description of a situation of the domain. An approximated joint probability distribution for \(X\) and \(Y\) is referred to by us as a context in which each possible situation-description has some probability of occurrence. In the following discussion we use the term domain to refer to an informally specified application area of problem solving and reasoning and by the term context we refer to a precise collection of past cases pertaining to some of the situation descriptions that can arise in the domain.

An instance of probabilistic reasoning is always performed within a particular context of a domain. This context is characterized by an approximated joint probability distribution that reflects knowledge about a subset of some possible situations that can possibly arise in a domain. This approximation is constructed from a database of some recorded cases from the domain. The scope of a particular context, is limited by the selection of the cases that are included in the database to approximate the joint probability distribution. All the available knowledge of a domain is generally too huge to be modeled by a single context. Also, for reasoning about a particular situation of a domain we do not want to include in the context modeling the situation all the information that may belong to the domain but has nothing to do with the particular situation being modeled. We would like to include in the context only those parts of the domain knowledge that may be relevant to the situation being reasoned about, and the relevance is determined by their being able to contribute towards making the desired types of inferences in the hypothesized models. Formally we can state our notion of relevance as follows:

- An attribute or a probabilistic dependency learnt from a particular context of a domain is relevant for hypothesizing a new arbitrary context only if it helps derive the desired type of inference in the hypothesized context.

The presence of those attributes and dependencies that are not relevant to the situation not only make the task of reasoning and handling knowledge unwieldy but also adversely affect the inferencing and the derived inferences.

The issues that we address below aim at constructing small and relevant contexts for modeling a particular situation and, specifically, are the following: 1. The various components of a domain's probabilistic knowledge; 2. The specification of relevance criteria for a decision making situation; and 3. the hypothesization of those consistent and minimal models that are also optimal from the perspective of the specified relevance criterion.

2 Components of Probabilistic Knowledge

One fundamental way in which our view of the probabilistic knowledge and reasoning differs from the traditional probabilistic reasoning methods is as follows. The traditional methods are focussed towards reason-
ing within a given context, that is, a given joint probability distribution. For a domain \( D \) it is assumed that a single database of cases is available and this database is used to approximate the joint probability distribution \(<H,PD>\) for the complete domain where, a single set of conditional probability dependency relationships, the set \( P_D \), characterizes the invariant relationships of the domain and the set of attributes is represented by \( H \).

In our view the available probabilistic knowledge for a domain is structured in one of the following two ways:

1. No single database for complete domain is available but a number of contextual databases, each corresponding to a specialized context of the domain, are available. The \( i^{th} \) of these contextual databases is represented by \(<T_i,P_i>\). It may not be possible to construct the complete probability distribution for the domain from these marginal component contextual databases. For example, in the domain of lung-diseases databases corresponding to different age-groups, geographical units, economic/racial subgroups, times of year etc. may be available. Here, each database represents a particular specialized context of the complete domain of lung diseases.

2. A single database of cases is available for some context. One can use some domain knowledge and split the database into entropy minimizing partitions such that each partition is taken as a specialized context of the domain. The database corresponding to the \( i^{th} \) partition can be compacted and represented as \(<T_i,P_i>\). If desired, the smaller contextual distributions can be recombined to obtain \(<H,PD>\) by retaining the relevant information during partitioning.

In both cases the domain knowledge is represented by \( n \) contextual databases, \(<T_1,P_1>,<T_2,P_2>\ldots\ldots<T_n,P_n>\) where the distribution \(<T_i,P_i>\) represents the \( i^{th} \) context \( S_i \) of the domain. It is possible that an attribute \( h \in H \) may appear in more than one context of the domain. In such a case there may exist a dependency of the type \( P[h|a.\text{subset.of.}\{S_i\}] \) in the set \( P_i \) of every context \( S_i \) that includes \( h \).

The various \( P_i \)'s represent the probabilistic dependency relationships that are considered the invariants within their respective contexts. If the causality knowledge available for a context is used in conjunction with the database then it is possible to ensure that each dependency \( P[a|\text{some}\text{.}\text{other}\text{.}\text{attributes} (S)] \) is such that for every attribute \( a \) the conditioning attributes are only those that can possibly causally affect the attribute \( a \). The dependencies in each set \( P_i \) can then be seen as the causal dependencies for the context.

We view each contextual joint distribution as a source of knowledge from which, using the available causal knowledge of the context, we can learn the invariant probabilistic causal dependencies for the context. An attributes which appears in a number of contextual databases may, therefore, be a part of one such dependency in each context in which it occurs.

The contextual causal dependencies learnt from individual contexts are reflective of the underlying causal processes for the context and therefore are viewed as possible causal building blocks from which other, possibly not yet encountered, contexts may be constructed. The causal dependencies learnt from a database containing cases from a number of contexts can be seen as weighted accumulations of the causal dependencies learnt from individual contextual databases.

3 Building a Hypothetical context

The model hypothesizing agent considers each invariant causal dependency, learnt from a particular contextual database, as a basic entity of the domain knowledge. He does not perform the traditional probabilistic inference in each context but seeks to build new contexts by putting together attributes and causal dependency relationships learnt from other contexts of the same database. The invariant causal dependencies of each context, therefore, are the basic units which the agent considers as representatives of the contexts' causal phenomena that can be used as building blocks for constructing the descriptions of other hypothetical contexts.

A hypothesized context, therefore, is constructed by selecting an appropriate subset of domain attributes and an appropriate subset of contextual causal dependencies from all the known contexts of the domain. A number of consistency and minimality constraints, however, must be satisfied by the chosen subsets. We have discussed one set of such criteria in [2].

When we mix and match the contextual dependencies learnt from disparate contexts to hypothesize new contexts we don't have enough information to determine the relative probability of occurrence of this new concocted context. This is because the marginals represented by the contextual databases are not sufficient to construct the complete joint distribution from which the probability of occurrence of a particular context may be determined. As long as the criterion for evaluating the relevance of a hypothesized context does not require the probabilities of occurrence of various scenarios or smaller contexts, the
above shortcoming is not of much significance.

The dependencies for each context may be learnt by constructing an equivalent Bayesian network [5], a Chow tree [3], or minimum entropy partitions of the database as discussed in [1].

We define a relation \( D \) for all the known contexts of a domain, that is, including information from all its contextual databases, such that

\[
D \subseteq \bigcup_{i=1}^{n-1} \{H^i \times H\} \text{ where } H^i = H \times H \times \ldots \times H, \quad \text{i times}
\]

where \( H \) is the set of all the attributes in all the known contextual databases, \( \text{domain}(D) = \bigcup_{i=1}^{n-1} \{H^i\} \) and \( \text{range}(D) = H \). This relation has the same form as the set \( P \) of dependencies of a joint probability distribution. The difference is that \( D \) contains all the dependencies learnt from all the contextual databases or the qualitative knowledge of the domain and it does not necessarily constitute a consistent description of a unique joint probability distribution.

4 Relevance Criteria

The relevance of a hypothesized context to the decision making situation may be evaluated in a number of different ways, depending on the objectives of the decision maker. Some possible candidates for the criterion are:

1. The probability of occurrence for some unobserved event \( d \) in the hypothesized context.
2. The ability of the hypothesized context to explain all the observed events of the situation.
3. The size of the hypothesized context in terms of the attributes and dependencies included in it.

The first of these criteria has been used by us in [2] for constructing those hypothetical situation models in whose contexts the probability of occurrence of some unobserved event \( d \) is the largest among all possible hypothetical models of the situation.

5 Structure of a Relevant Context

Since the description of a context is the same as a representation of a joint probability distribution, we can see a Bayesian network as representing a context. The task of hypothesising a context is, therefore, the same as the task of constructing a particular Bayesian network by selecting appropriate subsets from the sets of nodes and probabilistic dependencies.

Every choice made from the subsets of domain attributes and causal dependencies does not constitute a valid description of a hypothetical context in the domain. Each choice must satisfy the following constraints in order to qualify as an intuitively valid model for the situation.

5.1 A Sufficiently Large Context

A hypothesized context should be sufficiently large in the sense that it should:

1. Include the attribute \( d \) corresponding to the event of interest.
2. Include the observed attributes that have been observed in the situation being modeled.
3. Include a subset \( R \) of the dependency relation \( D \) such that their associated conditional probability functions define a complete joint probability distribution.
4. Be an Explanation for the occurrence of the desired event \( d \) and the observed events.

There are many different ways in which a particular context can be defined to be an explanation of the desired and the observed events. Consider an example of a patient with the symptoms Runny nose, sneezes, ear-ache, head-ache, and skin rashes. The desired event \( d \) for this case is fever. A hypothesized context can be of the following two types:

1. It has the shape of a connected Bayesian network and all the five observed symptoms and the desired event are at the leaves of this network. This may be the case when a set of causes is causally affecting each of the observed and desired symptoms.
2. It has the shape of two disjoint Bayesian networks one of which includes the symptoms runny-nose and sneezes and the desired event fever as the leaf nodes and the other network includes the remaining three observed symptoms at the leaf level.

The explanation formed by first Bayesian network connects each observed event with the desired event in the hypothesized context. In the second explanation the three symptoms of the second network have a causal explanation which is completely isolated from the causal explanation for the two observed and the desired event. The desired event is connected only to the two observed events and the other three observed events have been hypothesised as irrelevant for inferences about the desired event.
Accepting only the first kind as valid explanations imposes unrealistic constraints on the hypothesised contexts. They must anyhow try to relate each observed event to the desired event. The second type include the vacuous contexts in which the desired event is not connected to any observed event. Our hypothesising agent works to create only those hypothetical contexts in which some specified inference about the desired event $d$ can be made. Therefore, the hypothesising agent connects only those observed events to the desired event which lead towards the specified inference. If an observed event, when connected to the desired event, does not contribute to the specified inference, an alternative explanation for the event is sought totally disconnected to the desired event.

### 5.2 A Consistent Context

The notion of consistency of a hypothesised context is derived from the perspective of a context being the same as a probability distribution. That is, a context $< T, R >$ is consistent only if the dependencies in $R$ completely describe some unique joint probability distribution for the attributes in $T$. This consistency condition ensures that the resulting context is neither under-specified nor over-specified.

### 5.3 The Minimal Context

A hypothesised context $< T, R >$ is considered minimal if the following are true:

1. Removal of any dependency $r \in R$ from the context would disrupt a path between either attribute $d$ and an observed attribute $e_i \in E$, or between observed attributes $e_i \in E$ and $e_j \in E$.

2. Every $t \in T$ is included in at least one $r \in R$.

### 6 Constructing Relevant Contexts

The computational task of constructing the optimally relevant contexts has been formulated by us [2] as a state-space search guided by an admissible heuristic function. The AI search algorithm $A^*$ can generate the relevant contexts in the decreasing order of their relevance. The criteria of consistency and minimaliy can be enforced while building the partial contexts in the search process. Also, heuristic functions can be built for various specified relevance criteria. These criteria generally seek the upper bound of the relevance evaluation for each partially constructed context description. Each partial context description is enhanced by including in it more causal dependencies at each stage of the search process.

### 7 Conclusion

Intelligent systems have tended to use the probabilistic knowledge of a domain for probabilistic reasoning which is performed in the spirit of inductive inference. The task of constructing only the relevant from among all the conceivable contexts and scenarios of a domain requires that we somewhat move away from only the most probable, the most likely, or the least cost focus and instead view the most relevant as the objective in seeking contexts for probabilistic inference. We have presented in this paper a formalism for using the probabilistic knowledge available in a domain, by methods different from those of traditional probabilistic reasoning. We have outlined the nature of sufficient, consistent, and minimal hypothetical contexts that can be conceived by a model building agent and have briefly outlined an AI search based computational formalism for arriving at the interesting relevant hypotheses.

### References


