An Ontology-based Bayesian Network Approach for Representing Uncertainty in Clinical Practice Guidelines

Hai-Tao Zheng, Bo-Yeong Kang and Hong-Gee Kim
Biomedical Knowledge Engineering Laboratory
Seoul National University, South Korea

ISWC 2007 Workshop on Uncertainty Reasoning for the Semantic Web, November 12th, 2007, Busan, Korea

Agenda

- Motivation
- Framework
- Encoding Uncertainty into a CPG Ontology
- Algorithm of BN Inference
- A Scenario Validation
- Conclusion
- Future work

Motivation

- Uncertainty management is required not only in the process of following practices guidelines, but also in an earlier phase of selecting which practice guidelines might be applicable to a given patient. For this type of application, the ability to deal with uncertain data and quantify the uncertainty in order to perform inference on missing value is critical.

- Most guidelines are designed in such way that a clinician should not proceed unless there is no uncertainty about any data item. This expectation is unrealistic.

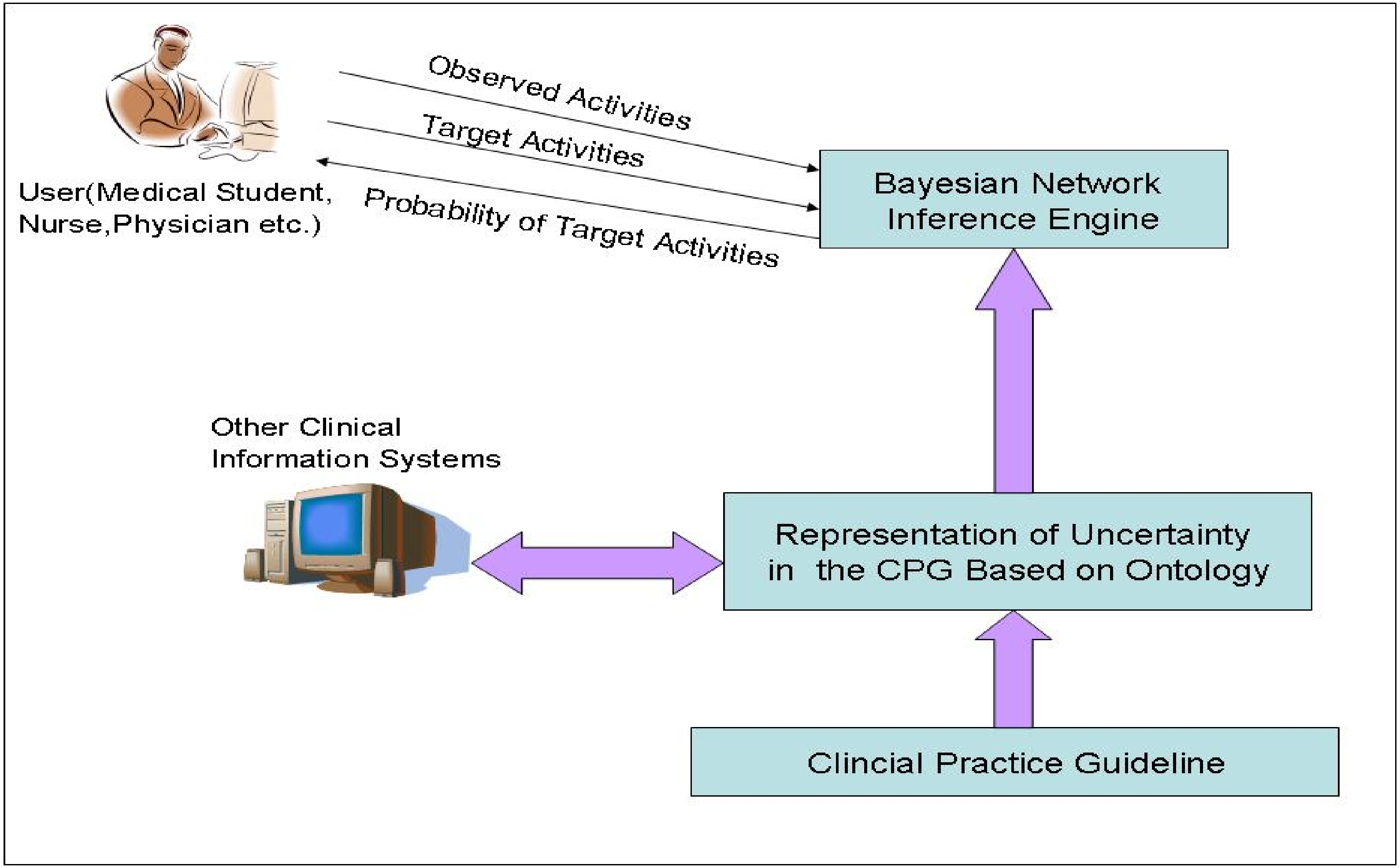
- Three sources of uncertain data have been identified: 1.data stemming from an unreliable sources; 2.data not obtainable; and 3.data not yet collected. [L. Ohno-Machado 2000]

Motivation

- The guideline should be able to proceed as long as principled inference could be made about the required data items.
  - For example, a guideline that requires that risk factors for heart disease be assessed, including risk of diabetes, may need to proceed even if the information on this item is uncertain.

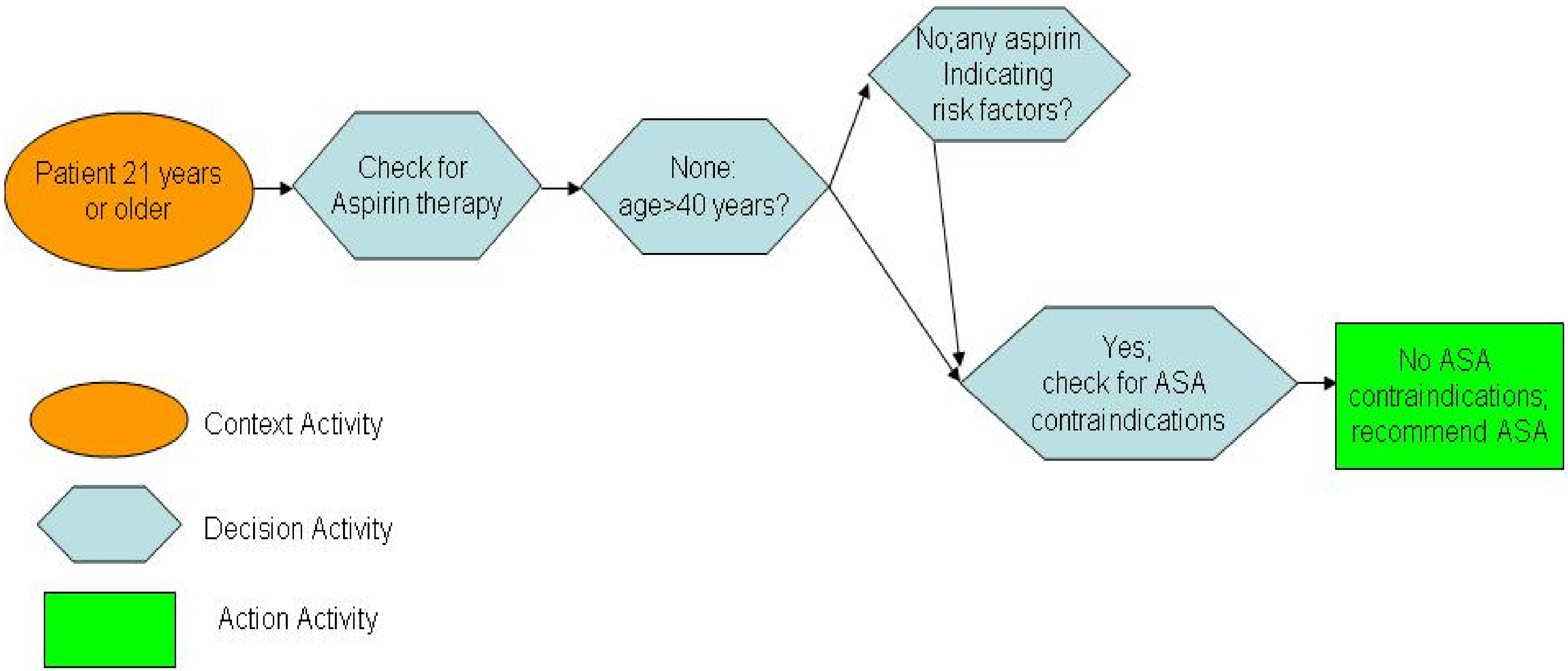
- A machine readable representation of uncertainty has to be filled the needs of the model that will be applied to perform principled inference.

Framework



Observed Activities
Target Activities
Probability of Target Activities
User(Medical Student,
Nurse,Physician etc.)
Bayesian Network
Inference Engine
Other Clinical
Information Systems
Representation of Uncertainty
in the CPG Based on Ontology
Clincial Practice Guideline

Encoding Uncertainty into a CPG Ontology

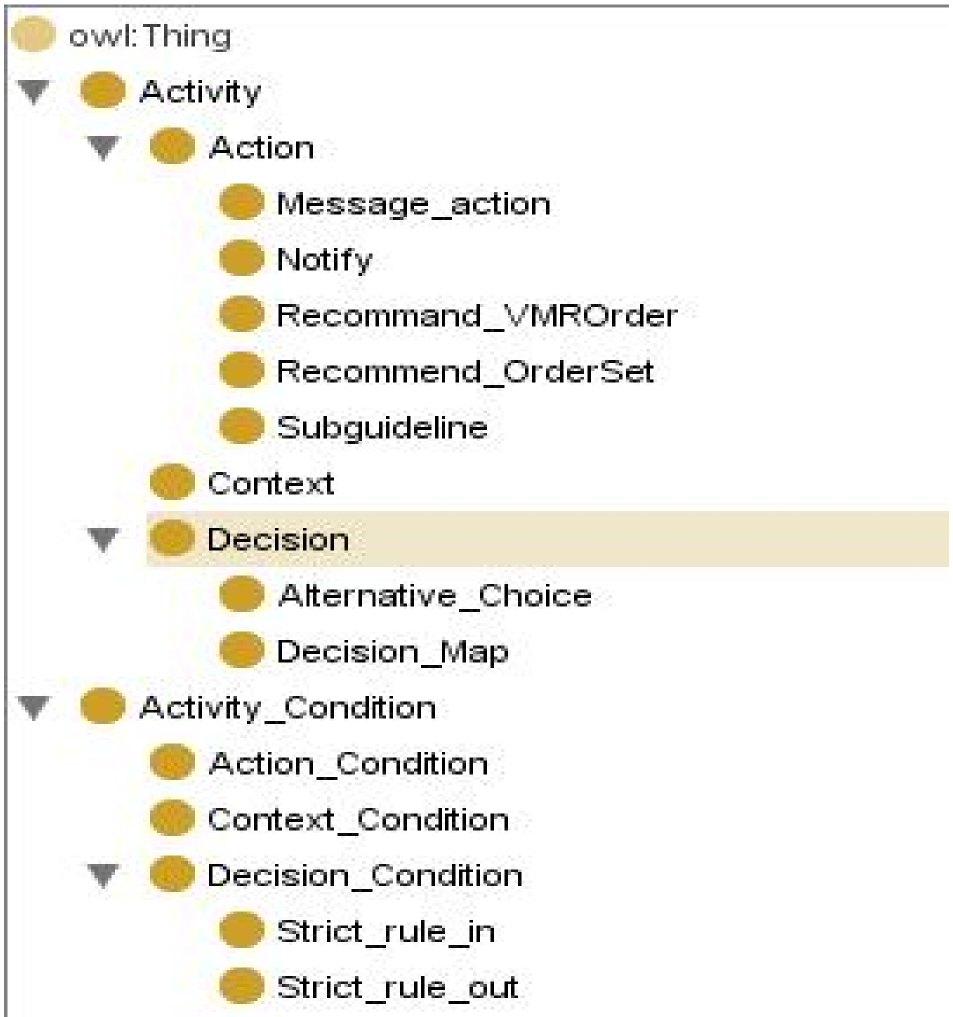


Clinical practice guideline of aspirin therapy for diabetic patients (ASA means aspirin therapy)

Three Kinds of Activities in CPGs

- **Context activity**
  - Each activity graph segment within a guideline begins with a context activity node that serves as a control point in guideline execution by specifying the clinical context for that segment.
- **Decision activity**
  - A decision activity node in the SAGE guideline model represents clinical decision logic by listing alternatives (typically subsequent action activity nodes), and specifying the criteria that need to be met to reach those nodes.
- **Action activity**
  - An action activity node encapsulates a set of work items that must be performed by either a computer system or persons.
- In CPGs, activities may include internal conditions that restrict the ir execution.

Classes in the CPG ontology



BiKE Biomedical Knowledge Engineering Laboratory
Seoul National Universtiy

A CPG ontology with Uncertainty Features

Definition 1. *(CPG Ontology) CPG Ontology* $O := \{C, I, Ps, cinst\}$, *with an activity class set* C, *an activity instance set* I, *a property set* Ps, *and an activity class instantiation function* $cinst : C \rightarrow 2^I$.

Definition 2. *(Properties for uncertainty representation) Property Set* $Ps :=$ $\{cause, hasCondition, hasState, isObserved, hasPriorProValue, hasCondiProValue\}$,

has a property function $cause : I \rightarrow I$, *a property function* $hasCondition : I \rightarrow I$, *a property function* $hasState : I \rightarrow Boolean$, *a property function* $isObserved: I \rightarrow Boolean$, *a property function* $hasPriorProValue: I \rightarrow Float$, *and a property function* $hasCondiProValue: I \rightarrow Float$.

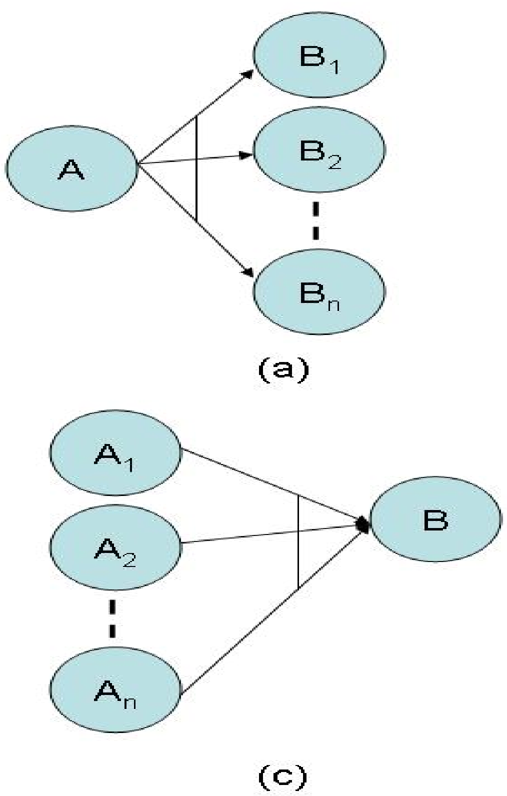
A Fragment of CPG Ontology

```
<Context rdf:ID="Patient_21_yo_or_older">
    <hasPriorProValue
    rdf:datatype="http://www.w3.org/2001/XMLSchema#float"
    >0.5</hasPriorProValue>
    <hasState
    rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean"
    >true</hasState>
    <cause rdf:resource="#Check_for_Aspirin_therapy"/>
</Context>
```

BiKE Biomedical Knowledge Engineering Laboratory
Seoul National Universtiy

Construction of CPTs in BN

- **We apply the max entropy assumption to learn the parameters.**



Construction of CPTs in BN

A causes the set of activities $\{B_1, B_2, ..., B_n\}$ simultaneously, the conditional probability $P(B_i|A) = 1.0, (i = 1, ..., n)$ (Fig. 4(a)); when an activity A causes one of the activities $\{B_1, B_2, ..., B_n\}$, the conditional probability $P(B_i|A) = 1.0/n, (i = 1, ..., n)$ (Fig. 4(b)); when a set of activities $\{A_1, A_2, ..., A_n\}$ cause activity B together, then $P(B|A_1, A_2, ..., A_n) = 1.0$ (Fig. 4(c)); when one of the activities $\{A_1, A_2, ..., A_n\}$ can cause activity B, then $P(B|\overline{A_1, A_2, ..., A_n}) = 0.0$ (Fig. 4(d)).

Algorithm of Bayesian Network Inference

- Variable elimination algorithm

$$P(\mathbf{X}) = \prod_i (P(X_i | pa(X_i))).$$

$$P(\mathbf{X}_q | E) = \frac{P(\mathbf{X}_q, E)}{P(E)} = \frac{\sum_{\mathbf{X} \backslash \{\mathbf{X}_q, \mathbf{X}_E\}} P(\mathbf{X})}{\sum_{\mathbf{X} \backslash \mathbf{X}_E} P(\mathbf{X})}$$

Algorithm of Bayesian Network Inference

Variable elimination algorithm [F. G. Cozman 2000]:

1. Generate an ordering for the N requisite, non-observed, non-query variables.
2. Place all network densities in a pool of densities.
3. For i from 1 to N:

(a) Create a data structure B_i, called a *bucket*, containing: the variable,called the *bucket variable*; all densities that contain the bucket variable, called the *bucket densities*;

(b) Multiply the densities in B_i. Store the resulting unnormalized density in B_i; the density is called B_i's *cluster*.

(c) Sum out X_i from B_i's cluster. Store the resulting unnormalized density in B_i's; the density is called B_i's *separator*.

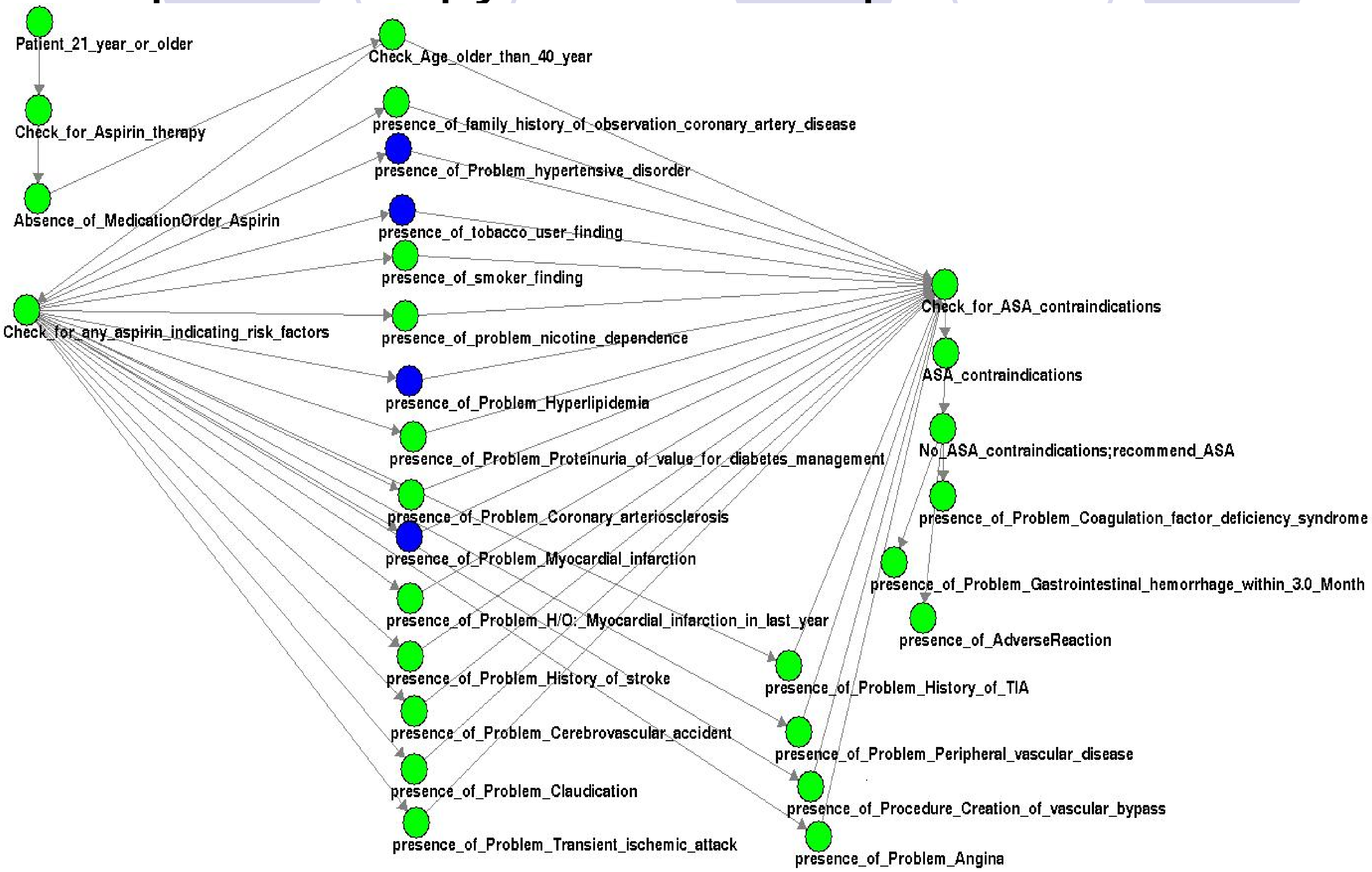
(d) Place the bucket separator in the density pool.

4. At the end of the process,collect the densities that contain the query variables in a bucket B_q. Multiply the densities in B_q together and normalize the result.

A Scenario Validation

- Scenario: a user (medical student, nurse or physician etc.) is trying to apply aspirin therapy for a diabetic patient using the diabetes CPG. When he/she tries to check the aspirin risk factors, he/she can get a few observed evidence, such as observations of hypertensive disorder, tobacco user finding, hyperlipidemia, and myocardial infarction. In this case, the user wants to evaluate target activities that he is concerned about in this CPG. In this way, he hopes the results can help him understand the effect of the observed evidence on the target activities during the whole clinical process.

An ontology based Bayesian network of aspirin therapy for diabetic patients



A Scenario Validation

$$P(\mathbf{X}_q|E)=\frac{P(\mathbf{X}_q,E)}{P(E)}=\frac{\sum_{\mathbf{X}\backslash\{\mathbf{X}_q,\mathbf{X}_E\}}P(\mathbf{X})}{\sum_{\mathbf{X}\backslash\mathbf{X}_E}P(\mathbf{X})}=0.775$$

where $\mathbf{X}_q$ ={ "No ASA contraindications; recommend ASA" }, and E={ "presence of problem hypertensive disorder" = $false$, "presence of problem myocardial infarction" = $false$, "presence of tobacco user finding" = $false$, "presence of problem hyperlipidemia" = $false$ }.

A Scenario Validation

$$P(\mathbf{X}_q|E) = \frac{P(\mathbf{X}_q, E)}{P(E)} = 0.6425$$

where $\mathbf{X}_q$={"presence of problem coagulation factor deficiency syndrome"} and E is the same as above case.

Conclusion

- In this paper, we contribute to present an ontology based BN approach to represent the uncertainty in CPGs.
- With this uncertain representation in ontology, computers can:
  - (1) calculate the uncertain degree of target activities in CPGs;
  - (2) remind users the missing important data or event items, which should be observed in the clinical process;
  - (3) simulate the clinical process under the uncertain situation, which can be applied to the e-learning systems in medical schools.

Future Work

- We will develop a more sophisticated learning algorithm based on experts opinions to construct the CPTs of BN.
- We will carry out a more comprehensive experiment to evaluate our approach.
- We will combine our approach with a real CIS (Clinical Information System) environment and apply uncertain clinical data to our application.

Question?

- Thank you