A Probabilistic Diagnostic Aid System on Internet, Using Data From a Large Medical Knowledge Base.

MIREILLE CLERET, FRANCK LE DUFF, ANITA BURGUN, PIERRE LE BEUX

LABORATOIRE PROTIMES, FACULTE DE MEDECINE - RENNES I
2 RUE HENRI LE GUILLOUX 35033 RENNES CEDEX FRANCE

Abstract DIAMED is a system to assist the physicians in the diagnostic process. It uses probabilistic networks as knowledge representation. By applying Bayesian methods it takes into account uncertainties of the facts in the resolution of clinical cases. The proposed model re-uses knowledge contained in an existing knowledge base (ADM) and orders them according to a semiological and pathological diagram.

MATERIAL AND METHOD

The French knowledge base ADM developed in Rennes University Hospital Center covers all medical specialties. This knowledge base contains more than 15,000 diseases and includes 110,500 entities. The diagnostic aid function of ADM gives full suggestions but the results are too noisy. Therefore, we propose a model using Bayesian probability theory. This model is a sectioned graph with several sub networks. The first level of graph relates signs of the same semantic kind and put them in the different semiological axis. For instance "constant thoracic pain" is included in the "thoracic pain" diagram in the "rhythm" axis. These signs-nodes are collected in semiological networks defined according to the type of examination (clinical, radiological...) and the topography target (heart, lungs...). The syndrome network restores the syndromes and allows to take in to account the signs belonging to different contexts. They are not independent when they result from the same pathological process. The pathological network links syndromes nodes to diagnosis nodes. This graph is the most complex of the model.

The propagation of evidence inside the sectioned network allows to solve clinical cases.

RESULTS

An interface of DIAMED developed on a Web server (http://www.med.univ-rennes1.fr/stat/) assists the experts of each medical specialty to update and validate the knowledge. To reduce the reformulation time, automated data processing is implemented: Constitution of lexicons starting from the existing dictionaries of the ADM system, selection of findings or diseases and updating. The lexicons are used to select all the signs relating to a particular sub-network and to build semiological axis. From a group of signs, the system selects the syndromes and builds semiological networks and the syndrome network. From these syndromes, DIAMED selects the concerned pathologies then successively the pathologies describing the previous ones. There also, the system has automatic process for detection of double arcs. The probabilities associated to nodes are computed according to frequencies and values corresponding to the accuracy of the findings. Thus, DIAMED takes into account the difficulties related to the availability of expert physicians time and to their knowledge of computer tools.

DISCUSSION

The proposed model is able to extract useful data from the ADM knowledge base in a diagnostic aid process and to arrange them in a semiologically ordered diagram. To decrease the amount of computations, the model contains simplifying hypotheses:

- Signs are independent from one another except those belonging to the same semiological network.
- Pathologies and findings are binary nodes.

Some restrictive hypotheses of other systems have not been used: independence of diseases or mutual exclusivity of diseases. The greatest interest of this model is its pseudo-segmentation that allows the propagation of information only to a part of the graph. It uses local network properties. Thus, the inference algorithms can use an exact method like that of Lauritzen and Spiegelhalter.

For physicians, this is an intuitive model: The reasoning is causal. The knowledge representation takes into account the semantic of the data, the hierarchies of the signs and the semiological mapping. Finally, DIAMED has the general power of probabilistic networks that is the independence of the knowledge base and of the inference algorithms, the possibility of using methods of learning the structures and parameters, and the control of base consistency.

REFERENCES
