Cross-Domain Probabilistic Inference: A Decision Support System for Dermatology and Rheumatology

Chang Ying-Jui[a,b] MD MSc, Li Yu-Chuan[b,c] MD PhD, Lee Chyou-Shen[d] MD, Hong-Sheng Shiue[b] MD, Yeh Min-Li[b,e] RN, MSN

[a]*Department of Dermatology, Far Eastern Memorial Hospital, Taipei, Taiwan*
[b]*Graduate Institute of Medical Informatics and* [c]*Department of Dermatology, School of Medicine, Taipei Medical University, Taipei, Taiwan*
[d]*Department of Internal Medicine, Division of Rheumatology, Mackay Memorial Hospital, Taipei, Taiwan*
[e]*Department of Nursing, Oriental Institute of Technology, Taipei, Taiwan*

Introduction

Non-infectious generalized blistering diseases (GBD) and autoimmune diseases (AID) are not rare in the dermatology and rheumatology fields. They share some common clinical findings and laboratory test results, which yield certain degree of uncertainty. Dermatologists and rheumatologists often cope with patients in multi-domains. We used previous well-established "Probabilistic Dermatopathological Clinical Diagnostic Decision Support System (CDSS)[1]" as mainframe, which contained a knowledge base (KB) for GBD; research group built up a new KB for AID. By mean of the mathematical formulation called "cross-domain Bayesian formulation", values of Apriori, TPR and FPR of the two KBs could be transferred in the different domains. Cross-domain CDSS was proved to be available and useful.

Method and Material

Two main parts are needed in the process of constructing a cross-domain decision support system:

(1) Knowledge presentation and system shell
(2) Cross-domain probabilistic inference

Knowledge presentation and system shell

Knowledge representation is the core of a decision support system which including methodology of knowledge engineering, structure of knowledge, and algorithm of inference engine. Probabilistic inference uses probability to present the uncertainty of a knowledge field; it also uses mathematical formula to calculate the inference result. The most popular one is Bayesian formulation[2].

Inference engine, user interface, and other tools in maintaining KB make up the "system shell". Our inference engine was built based upon multi-membership Bayesian formulation[3] of knowledge representation. We used a Web-based interactive interface due to its friendly and graphical user interface. Users can access this system via a WWW browser without geographic or horary limitations as long as they are connected to the Internet.

In the process of the construction of CDSS, knowledge acquisition is the most time- and human-consumption step. The process of transferring medical knowledge to KB which could be utilized by computer is called "knowledge engineering". In this process, major human resources involved are GBD/AID experts, coordinators, knowledge engineers, and programming engineers. In constructing the medical KB, a Bayesian disease frame was constructed to represent each disease in the GBD/AID domain. As each frame was built, domain experts decide which findings are pertinent to that disease. Apriori of each disease, true positive rate (TPR) and false positive rate (FPR) of each finding were obtained for these frames from literature search, healthcare database statistics and experts' estimates[2,4,5].

Cross-domain probabilistic inference

It is different that the Aprioi of a certain disease in different domain, and so is the TPR and FPR of a certain diagnostic criteria of clinical finding. The vo domains is quite important. We h medical domain "A" and "B", and ne of the diagnostic criteria. The as figure 1.

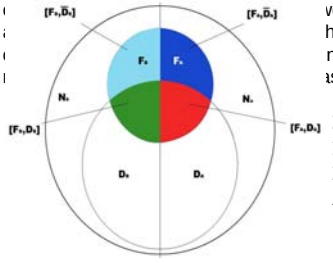

Figure 1. The relationship of the disease and medical domains.
D_A: disease in domain A; D_B: disease in domain B; F_A: finding in domain A; F_B: finding in domain B
N_A: total patient number of domain A; N_B: total patient number of domain B; $P[F_A|D_A]=a$, $P[F_B|D_B]=b$, $P[\overline{F}_A \mid \overline{D}_A]=\alpha$, $P[\overline{F}_B \mid \overline{D}_B]=\beta$

$$P[D]=P[D_{A\&B}]=\frac{N_a P[D_A]+N_b P[D_B]}{N_a+N_b}$$

$$P[F,D]=P[F_A,D_A]+P[F_B,D_B]=a\times P[D_A]+b\times P[D_B]$$

According to the assumption, the probability of disease D is $P[D]$, while

$$P[F,\overline{D}]=P[F_A,\overline{D}_A]+P[F_B,\overline{D}_B]=\alpha\times(1-P[D_A])+\beta\times(1-P[D_B])$$

$$TPR=P[F\mid D]=\frac{P[F,D]}{P[D]}=\frac{(N_a+N_b)(aP[D_A]+bP[D_B])}{N_a P[D_A]+N_b P[D_B]}$$

$$FPR=P[F\mid \overline{D}]=\frac{P[F,\overline{D}]}{P[\overline{D}]}=\frac{(N_a+N_b)(\alpha(1-P[D_A])+\beta(1-P[D_B]))}{(N_a+N_b)-(N_a P[D_A]+N_b P[D_B])}$$

Probability of D if F is present

$$P[D\mid F]=\frac{P[D]\times TPR}{P[D]\times TPR+(1-P[D])\times FPR}=\frac{aP[D_A]+bP[D_B]}{(a-\alpha)P[D_A]+(b-\beta)P[D_B]+(\alpha+\beta)}$$

Probability of D is F is absent

$$P[D\mid \overline{F}]=\frac{P[D]\times(1-TPR)}{P[D]\times(1-TPR)+(1-P[D])\times(1-FPR)}=\frac{N_a P[D_A]+N_b P[D_B]-(N_a+N_b)(aP[D_A]+bP[D_B])}{(N_a+N_b)(1-(aP[D_A]+bP[D_B]))+(\alpha P[D_A]+\beta P[D_B])-(\alpha+\beta))}$$

By the similar deduction, we can get the general form of cross-domain Bayesian formulation as shown below:

$$P[D]=\frac{\sum_i^n N_i\times P[D_i]}{\sum_i^n N_i}$$

$$TPR=P[F\mid D]=\frac{\sum_i^n N_i\times\sum_i^n P[F_i\mid D_i]\times P[D_i]}{\sum_i^n N_i}$$

$$FPR=P[F\mid \overline{D}]=\frac{\sum_i^n N_i\times\sum_i^n P[F_i\mid \overline{D}_i]\times P[\overline{D}_i]}{\sum_i^n N_i-\sum_i^n N_i\times P[D_i]}$$

The symbol i refers to a medical domain.

We use the cross-domain Bayesian formulation as inference engine which programmer will incorporate into the core of the programming language.

Results

Two KBs (GBD & AID) were built:
11 disease frames, 90 findings, and 171 values of Aprioi, TPR and FPR for GBD.
6 disease frames, 98 findings and 78 values of Aprioi, TPR and FPR for AID.
Cross-domain Bayesian formulation was used to convert the different values between two domains.

GBD KB was proved to be available in the previous study. 20 cases were abstracted from case-report articles in respected rheumatological journals for AID KB validation. After calculation, 17 cases were given the correct diagnoses by the system. The consultation results of the remaining three cases were ranks the second rank of the possible diagnosis. The non-error rate was 85%(17/20). The average of probabilities assigned to the correct diagnosis was 66.3%.

The second step is to validate the cross-domain consultation system. 10 cases were abstracted from journals for testing. The findings were entered into the cross consultation section. At the same time, cases were also selected for individual KB consultation manually. After calculation, system showed the result of non-error rate was 90%(9/10). The average of probabilities assigned to the correct diagnosis was 64.76% (Table 1)

	Gold standard	Cross-domain Dis	%	GBD Dis	%	AID Dis	%
Case1	SLE	SLE	55.3			SLE	15.7
		PA	0.7			PA	0.7
Case2	SLE	SLE	55.3			SLE	15.7
		PA	3.2			PA	3.2
Case3	SLE	SLE	80.45			SLE	80.5
		PA	0.7			PA	0.7
Case4	SLE with edema	SLE	69.8			SLE	41.7
		BD	3.8			PA	0.7
Case5	SLE with Sjogren	SS	7.18			SS	7.2
		PA	0.75			PA	0.7
Case6	SLE with psyocsis	SLE	7.27			SLE	7.3
		PA	0.75			PA	0.7
Case7	BP with SLE	BP	92.58	BP	92.6		
		SLE	31.72	PDA	27.4		
Case8	PV with SLE	PV	99.77	PV	99.8		
		SLE	13.94	TAD	0.3		
Case9	Bullous DM	DM	99.78			DM	99.8
		BD	7.2			DM	<1
Case10	Erythema of Sjogren	SS	22.6			SS	22.6
		PA	3.6			PA	3.6

Table 1. The probabilities of cross-domain consultation. BP: Bullous Pemphigois; DM: Dermatomyositis; EBA: Epidermolysis Bullosa Acquisita; PM: Polymyositis; PV: Pemphigus Vulgaris; RA: Rheumatoid Arthritis; SLE: Systemic Lupus Erythematosus; SS: Sjogran's Syndrome; TAD: Transient Acantholytic Dermatosis

Conclusion and Discussion

We have developed a Web-based probabilistic inference engine and shell for CDSS that deals with uncertainty explicitly. We also engineered a KB for diagnosing GBD and AID that proved to be quite accurate when given cases from medical journals. This CDSS could aid physicians in differentiating rare disease groups such as GDB and thus help them make better diagnostic and treatment decisions. This knowledge-based system could also help medical students in learning to diagnose diseases when facing suspicious cases, though further evaluation is required.

To the authors' knowledge, this is the first cross-domain CDSS with a Web-based interface where KBs can be built and maintained on the Internet. It is of good availability and ease-of-use, and could be integrated into other clinical systems. This preliminary evaluation result also demonstrated that such a CDSS could be successfully implemented in a Bayesian formulation with a Web interface. We proposed this cross-domain probabilistic inference upon a Web-based CDSS for dermatology and rheumatology, and we believe it maybe very well be the first such probabilistic decision support system developed in the world.

Reference

1. http://ades.tmu.edu.tw/piew/
2. Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical decision making.* Butterworth-Heinemann, Boston. 1988.
3. Ben-Bassat M. Multimembership and multiperspective classification: introduction, applications, and a Bayesian model. *IEEE Trans Syst Man Cybern.* SMC-ID6: 331-6, 1980.
4. Warner HR, Sorenson DK, Bouhaddou O. *Knowledge Engineering in Health Informatics.* Springer-Verlag, New York, 1997.
5. Soc HC. Probability theory in the use of diagnostic tests. An introduction to critical study of the literature. *Ann Int Med.* 1986; 104(1): 60-6.