Dempster-Shafer 이론의 근사방법을 이용한 전문가 시스템의 불확정성 처리

이 종 만, 김 진 형

한국 과학 기술원 전산학과

요 약

전문가 시스템이 부정확하고 불충분한 자료를 이용하여 인간 전문가가 내리는 판단과 비슷한 추론을 하도록 하기위하여 불확정성의 문제에 관한 많은 연구가 있다. 이중 Dempster-Shafer(D-S) 방법은 다른 방법에서 찾아볼 수 없는 많은 장점이 있으나 지수적 복잡도 때문에 현실적 응용에 많은 장애가 되고 있다. 본 논문에서는 이 복잡도를 선형의 복잡도로 낮추기 위하여 기존의 방법에서 제기되었던 문제점 중에서 계산상의 오차를 줄이고 불확증 자료의 신뢰도 표현및 신뢰도 구간의 표현이 가능하도록 새로운 근사 방법을 제시하였다. 임의의 표본자료를 이용하여 분석한 결과 기존의 방법보다 오차가 적고 계산시 고려되는 집합수가 더욱 증가해도 이에 따른 오차의 감소율이 둔화되므로, 복잡도와 정확성을 고려할때 본 논문에서 제안된 방법이 가장 효율적인 방법임을 입증하였다.

I. 소 개

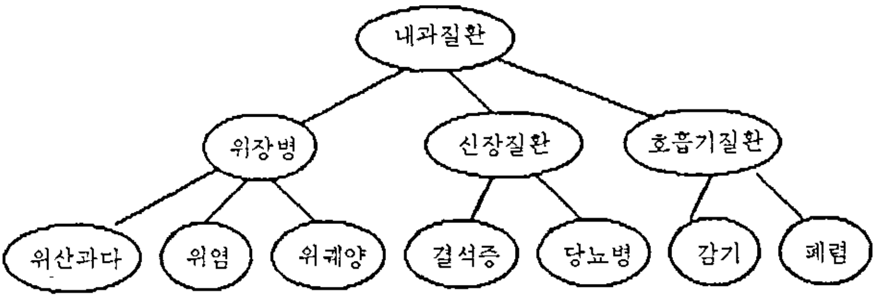
전문가 시스템이 여러 분야에 널리 사용되면서 주어진 자료가 비록 불완전하고 부정확하지만 이 자료를 이용하여 인간 전문가가 내리는 판단과 비슷한 결과를 얻을 수 있도록 하기 위한 불확정성의 문제가 더욱 중요시 되고 있다. 불확정성의 문제 해결을 위하여 사용되는 방법에는 여러가지가 있다. 첫째, 확률에 기초를 두고 있는 베이스 이론[1]은 기존의 여러 방법중 가장 정확한 결과를 얻을 수 있으나, 이 방법을 적용하기 위해서는 너무나 많은 자료를 요구하는 단점이 있다. 또하나의 방법은 내과 진료용 전문가 시스템인 MYCIN에 이용된 CF(Certainty Factor) 방법[2, 3]으로서, 개념이 비교적 간단하여 사용시 효율성은 좋으나 수학적 기반이 없는 단점이 있다.

베이스 이론 대신에 사용할 수 있는 Dempster-Shafer(D-S) 방법[4]은 최근에 많은 연구가 되고 있는 방법으로서 CF 방법보다는 효율성이 나쁘며 고려되는 전체 영역내의 집합들이 서로 독립적인 성격을 가져야 하는 전제조건이 필요하지만 신뢰도 구간의 표현및 고려되는 집합들에 대해 아직 모르는 부분에 대한 표현이 가능하다는 장점이 있다. 이 D-S 방법은 주어진 자료들을 결합시키는 과정에서 지수적 복잡도를 가지므로 실제 시스템에 이용할 수 있도록 이를 선형의 복잡도로 낮추기 위하여 근사 방법이 이용되며, 이에따라 오차가 발생하고,

D-S 이론의 가장 큰 장점인 신뢰도 구간및 불확증 신뢰도의 표현을 할 수 없게 되는 단점이 있다. 이 논문에서는 기존의 근사 방법보다 오차가 적으면서 신뢰도 구간의 표현및 불확증 신뢰도의 표현이 가능한 D-S 방법의 새로운 근사방법을 제시하고자 한다.

II. Dempster-Shafer(D-S) 이론의 소개

전문가 시스템의 추론기능은 고려되는 각 집합들에 대해 수집된 자료들을 종합하여 가능성 있는 각 집합에 대해 그 가능성의 정도를 나타내 주는 과정으로 볼 수 있다. 본 논문의 이해를 돕기 위하여 내과의사가 진료하는 질병들 중에서 위장병과 호흡기 질환및 신장 질환에 대해 예를 들어 설명하겠다. 위장병은 다시 위산과다, 위염및 위궤양으로 세분되며, 호흡기 질환은 감기와 폐렴으로, 신장 질환은 결석증과 당뇨병으로 각각 세분되며, <그림 1>이 이들 질병 사이의 연관관계를 설명하고 있다.



<그림 1> 내과 질환의 계층 구조

D-S 이론에서 위의 7가지 질병 {위산과다, 위염, 위궤양, 결석증, 당뇨병, 감기, 폐렴}을 식별되는 전체구조(Frame of Discernment)라 하며 Θ로 나타낸다. 이 Θ내의 각 원소들은 상호 배타적이며 소모적인 성격을 가지며 상위 계층의 질병은 하위 계층의 질병들을 포함한다 즉, 한가지 질병은 하위 계층의 다른 질병들의 집합으로 구성된다. 위 <그림 1>의 11가지 각 질병들에 대해 확증 자료와 함께 불확증 자료도 제공될 수 있다.

Θ의 각 부분 집합에 대한 고유의 자료는 기본적 확률 배당 함수(basic probability assignment, bpa)로서 나타내며, m으로 표시한다. Θ의 한 부분집합인 A에 대한 m(A)의 값은, A 자체에만 배당된 확신도 값을 나타내며, 이 값은 A를 구성하는 다른 부분집합의 확신도 값으로 세분되지 않는다. m(Θ)는 Θ에 배당된 확신도 값으로서 Θ의 부분집합들에 배당되고 아직 남아있는 확신도 값을 나타낸다. 즉, Θ의 부분집합인 A에만 값 s가 배당되어 있

다면 m(Θ)에는 1-s가 배당된다. 각 질병에 대한 신뢰도는 신뢰도(Belief) 함수 Bel로 표시되며 임의의 부분집합 A에 대한 Bel값은 A의 모든 부분집합들의 기본적 확률 배당함수(m)값을 합한 값으로 정의된다. 단일 요소로 구성된 집합에 대한 Bel값은 m값과 동일하며, Bel(Θ)는 항상 1인 특성을 가진다.

각 부분집합에 대해 확증과 불확증을 나타내는 자료들은 상호 결합되는 과정이 필요하다. 이를 위해 D-S 이론에서는 Θ의 부분집합이면서 서로 다른 관찰에 의해 수집된 집합들의 값들을 결합시켜 새로운 m값을 만드는 연산 방법으로서 ⊕로 표시되는 orthogonal sum을 제공하고 있다. Bel1, Bel2가 두개의 신뢰도 함수이며 m1, m2는 각각 이와 연관된 기본적 확률 배당 함수일때, Θ의 부분집합 X, Y의 orthogonal sum인 m1(X)⊕m2(Y)는, m1(X)와 m2(Y)의 각 값이 곱해진 결과가 X와 Y의 공통되는 부분집합에 배당된다. m1⊕m2(A)는 위와 같이 계산된 모든 결과들 중에서 부분집합이 A인 것들의 값을 모은 것이다. Bel1⊕Bel2(A)도 A의 모든 부분집합 B에 대한 m1⊕m2(B)의 값을 합하여 구할 수 있다. 이때 X와 Y의 공통되는 부분집합이 없으면, 1에서 이 값을 뺀 나머지 값으로서 전체를 재조정하여 다시 전체의 합이 1이 되도록 한다.

Bel(A)와 Bel(A^C)는 각각 부분집합 A에 대한 확증도와 불확증도를 나타낸다. 1-Bel(A^C)는 Bel(A)가 가질 수 있는 최대값을 나타낸다. Bel(A)가 가질 수 있는 값의 범위를 신뢰도 구간(Belief Interval)이라 하며 [Bel(A), 1-Bel(A^C)]로 표시한다. 이 구간의 폭은, 수집된 자료들을 모았을때 아직 부분집합 A에 대해 알려지지 않은 불확실성을 나타낸다.

III. Dempster-Shafer 방법의 적용 예

1. 단일 요소로만 구성된 집합에의 적용

위에서 언급했듯이 D-S 이론은 각 부분집합에 대한 불확증 신뢰도와 신뢰도 구간의 표현이 가능한 장점이 있다. n개로 구성된 집합에서 실제 부분집합들의 결합시에는 Θ의 모든 부분집합인 2^n개의 부분집합이 생성되므로 계산의 복잡도가 높아 실제로 이용하기에는 한계성이 있다. 2^n개의 부분집합들 중에서 상호 배타적인 단일요소로 구성된 집합에 대해서만 이의 확증 자료와 불확증 자료만을 고려하여 복잡도를 낮추는 방법이 Bornett에 의해 제안되었다 [5] 이 방법은,

(1) 각 부분집합에 대해 확증 자료와 불확증 자료들을 각각 결합시킨다.

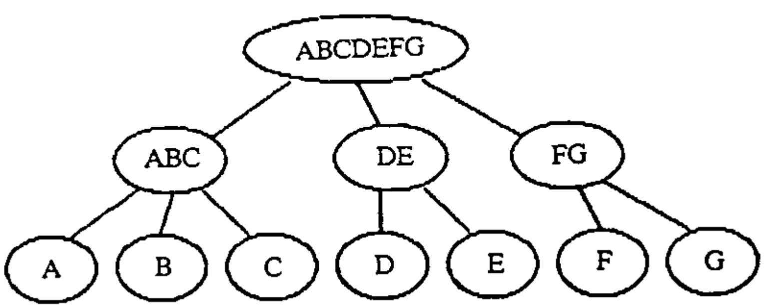
(2) 각 부분집합에 대해 위에서 구한 두가지 값들을 하나로 결합시킨다.

(3) 위의 결과로 생기는 n개를 결합시킨다.

이 방법을 이용하면 단일요소만으로 구성된 집합들은 선형의 복잡도로 계산이 가능하다.

2. 계층 구조로 구성된 집합에의 적용

단일 요소만으로 구성된 집합에 대해서만 고려한 위의 방법은 계층구조로 구성된 집합에 대해서는 적용될 수 없는 단점이 있다. 고려되는 각 질병들이 <그림 1>과 같이 계층구조로 구성되어 있을때 각 질병들에 대한 값은 임의의 계층에서도 입력되어 결합될 수 있어야 D-S 이론이 다양한 분야에 사용될 수 있다. 전문가 시스템에 사용되는 전문지식이 계층 구조로 표현될 수 있으면 D-S 방법이 이용될 수 있으나 이 방법이 가지고 있는 지수적 복잡도를 줄이기 위한 방법이 필요하다. 이러한 복잡도는 각 부분집합들을 결합시 불확증 자료들의 결합시에 발생되므로 불확증 자료들의 결합시 다음과 같은 근사 방법을 사용하여 계산의 복잡도를 줄이는 방법이 있다[6].



<그림 2> 내과 질환 계층구조의 부호적 표현

<그림 2>는 <그림 1>의 내과질환 계층구조의 부호적 표현이다. <그림 2>와 같은 각 집합들에 대해 관찰에 의해서 확증집합 {A, B, C, D, E, F, G, ABC, DE, FG, ABCDEFG}과 불확증 집합 {A^C, B^C, C^C, D^C, E^C, F^C, G^C, ABC^C, DE^C, FG^C, $ABCDEFG^C$}에 대해 각각 자료가 주어졌을때 이들을 결합시키는 방법은 다음과 같다.

(1) 위의 계층구조에서 주어진 11개의 각 부분집합에 대한 확증 자료와 불확증 자료들을 각각 결합시킨다.

(2) 고려되는 대상의 모든 부분집합에 대한 확증 자료들을 먼저 결합시킨다. 이때에는 이미 확증 집합에 속해 있는 부분집합들만이 이용되므로 쉽게 계산이 가능하다.

(3) 위의 결과에 고려되는 대상의 각 부분집합들의 불확증 자료들을 결합시킨다. 불확증

자료는 이에 상응하는 확증자료의 표현으로 바뀌어 결합되며, 새로 생성된 부분집합이 이미 주어진 확증집합에 없는 것이면 이에 가장 가까우면서 이것을 포함하는 확증집합으로 근사시켜 이곳에 계산된 값을 저장한다.

계층 구조를 가지는 집합들에 대해서도 이러한 근사 방법을 이용하면 선형의 복잡도로서 계산이 가능하다. 그러나 이에 따라 발생되는 문제점은 다음과 같다.

(1) 근사 계산에 따라 오차가 발생되며,

(2) 각 부분집합에 대한 불확증 자료는 이에 상응하는 확증집합으로 변형되어 계산되므로 모든 집합들에 대한 자료가 결합된후 각 집합에 대한 불확증 신뢰도를 알 수 없게 되며,

(3) 이에 따라 각 집합에 대한 신뢰도 구간의 표현도 불가능해 진다.

위의 3가지 단점들중 첫번째 단점인 오차를 줄이며, 둘째및 셋째의 단점을 제거하기 위한 새로운 근사방법을 다음절에서 제시하겠다.

IV. D-S 방법의 새로운 근사방법과 평가

1. D-S 방법의 새로운 근사 방법

계층구조로 구성된 각 부분집합들에 대해 초기에 주어지는 자료는, 각 부분집합에 대한 확증 자료와 불확증 자료가 주어진다. 앞절에서 설명한 근사방법에서는, 이 두가지 자료가 주어짐에도 불구하고 확증 자료에 대한 집합들만을 계산시 고려하여 불확증 자료는 이에 상응하는 확증 자료의 표현으로 근사시켜 계산에 이용함에 따라 앞절에서 설명한 3가지 단점이 발생하였다. 그러나, 자료가 정확히 주어진 자료들에 대해서는 계산시 해당되는 집합을 고려하여야 근사치의 이용에 따른 오차를 줄일 수 있고, 전체 자료의 결합후에도 각 집합에 대한 불확증 신뢰도 및 신뢰도 구간의 유지가 가능해진다. 이와같이 계산시 고려하는 집합의 갯수를 자료가 정확히 주어진 모든 집합들로 증가시킴에 따라 계산시 적용될 알고리즘은 다음과 같이 수정된다.

(1) 각 집합별로 확증 자료와 불확증 자료들을 각각 결합시킨다.

(2) 고려되는 영역내의 모든 집합들의 확증 자료들을 결합시킨다.

(3) 위의 결과에 모든 집합들의 불확증 자료들을 결합시킨다. 이때 새로 생성된 부분집합이 고려되는 전체 영역의 확증 집합과 불확증 집합에 없는 집합이면, 이에 가장 가

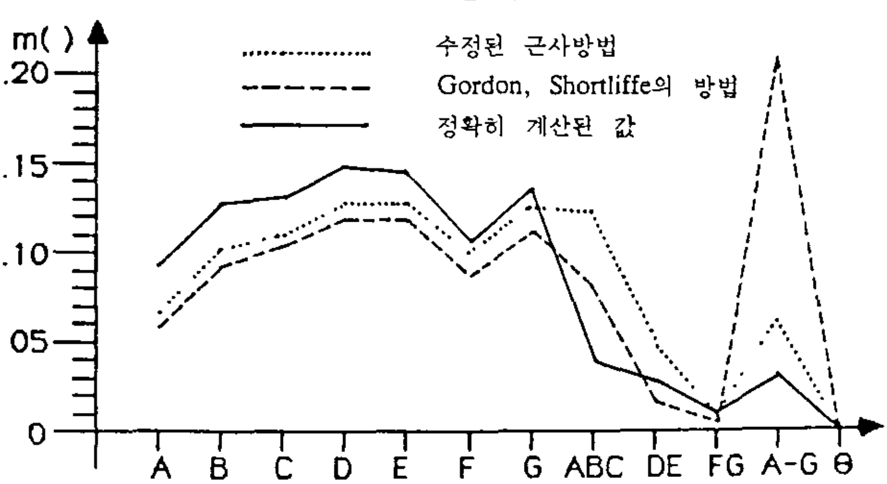
까우면서 이것을 포함하는 집합들을 구하여 이 집합들에 계산된 값을 균등히 배분한다. 즉, Gordon, shortliffe의 방법에서는 새로 생성된 부분집합을 확증 집합중의 부분집합에 근사시켰으나 본 논문에서 제시된 방법에서는 자료가 정확히 주어진 집합들인 확증 집합과 불확증 집합에 근사시켰다.

2. 새로운 근사 방법의 평가

일반적으로 계산시 고려되는 부분집합의 수가 증가하면 정확도는 향상되지만 복잡도도 따라서 증가한다. 본 논문에서 재안한 근사방법은 Gordon, Shortliffe의 방법보다 계산시 2배 많은 부분집합들을 고려한다. 이에 따라 n이 확증집합의 수일때 복잡도도 O(n)에서 O(2n)으로 증가하지만 실제 수행결과 CPU time은 1.5배만이 증가하였다. 그러나, 오차의 감소와 불확증 집합의 신뢰도및 신뢰도 구간의 표현등 3가지 장점을 얻을 수 있다.

부분집합	확증자료	불확증자료
A	0.50	0.62
B	0 60	0 61
C	0 70	0 72
D	0 55	0 54
E	0 58	0.58
ABC	0.52	0.64
DE	0 54	0 63
F	0 62	0 72
G	0.68	0 71
FG	0.51	0 64
ABCDEFG	0.52	0 58

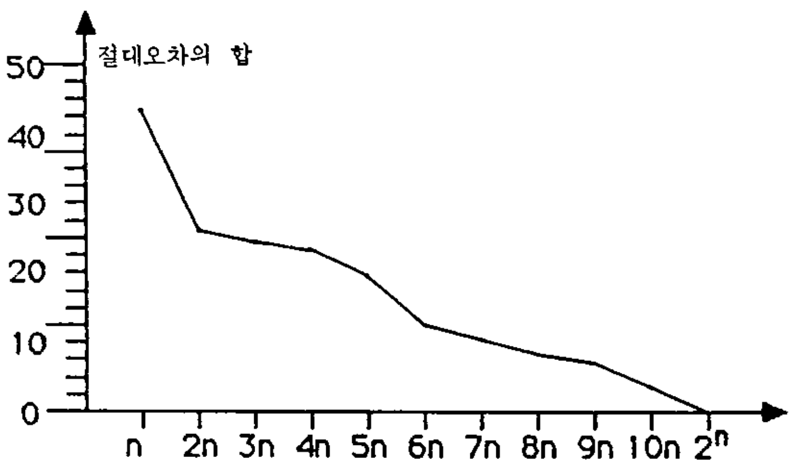
<표 1> 표본 자료



<그림 3> 표본 자료를 이용한 근사 방법의 결과 비교

위의 <그림 3>은 <표 1>의 표본자료를 이용하여 계산된 것으로서 수정된 근사방법에 의한 결과와 Gordon, Shortliffe의 방법에 의한 결과및 정확한 계산에 의한 결과를 나타낸다. <그림 2>의 계층구조를 고려하여 계산된 결과를 비교해보면, 하위 계층및 상위 계층에서는 오차가 줄었으나 중간 계층인 ABC, DE등에서는 오차가 오히려 증가하였다. 그 이유는, Gordon, Shortliffe의 방법에서는 하위 계층인 A, B, C, D, E, F, G등에 배당될 값이 상위 계층인 ABCDEFG에 근사되어 배당되었으나, 새로 제안된 방법에서는 이 하위 계층의 값들이 중간 계층인 ABC, DE등에 근사되어 배당되었기 때문이다. 즉, 계층구조의 계층이 많아지면, 새로운 근사방법에 의한 계산이 배당될 값의 상향화 현상을 감소시키므로 보다 정확한 계산이 가능해진다.

다음의 <그림 4>는 고려하는 집합들의 수를 확증집합의 수만큼씩 더 증가시켰을 경우에 정확한 결과에 대한 각 부분집합들의 절대오차의 합을 나타낸다. 여기에서 n은 확증집합의 갯수이다.



<그림 4> 고려되는 집합수의 증가에 따른 오차의 감소 경향

Gordon, Shortliffe의 방법은 계산시 확증집합만을 고려하기 때문에 고려되는 부분집합의 갯수가 n이며, 본 논문에서 제시한 근사방법은 확증집합과 불확증집합을 고려하기 때문에 2n에 해당된다 3n부터 10n까지는, 2^n개의 부분집합중 위의 2n개를 제외한 나머지 부분집합들 중에서 random number를 이용하여 임의로 n개씩 선택하여 계산에 이용하였다. 위의 <그림 4>에서 고려되는 집합수가 n에서 2n으로 증가하면 오차가 급격히 감소되었다. 그러나, 3n개 이상이 고려될 때에는 고려되는 부분집합의 수를 늘려도 오차의 감소현상이 둔화되므로, 복잡도와 효율성을 같이 고려한다면 본 논문에서 제시한 2n개인 경우가 가장 효율적임을 알 수

있다.

V. 결론

D-S 방법을 실제 적용하는데 있어서의 문제점은, 이 방법이 가지는 지수적 복잡도이다. 각 집합들이 계층구조로 구성되어 있을때 이 D-S 방법을 선형의 복잡도로서 효율적으로 이용할 수 있는 방법을 제시하였다. 이 방법은 자료가 정확히 주어진 부분집합들을 계산시에도 이용하여 생성된 부분집합들을 이에 근사시키는 방법으로서, 계산시 발생되는 오차를 줄일 수 있고 각 부분집합에 대한 불확증 신뢰도와 신뢰도 구간의 표현등이 가능한 장점을 가진다. 또한, 고려하는 집합수를 더욱 증가시켜도 증가하는 복잡도에 따라 오차의 감소가 적으므로, 복잡도와 정확성의 측면에서 이 방법이 효율적이라 할 수 있다. 이 방법은 관계되는 집합들이 서로 배타적이고 계층구조로 구성될 수 있다면, 진료와 진단및 분류 시스템등에 추론 방법으로서 유용하게 이용할 수 있다.

참 고 문 헌

[1] Duda, R., Hart, P., and Nilsson. N., Subjective Bayesian Methods for Rule-Based Inference Systems, in: *Proceedings 1976 National Computer Conference*, (AFIPS, 1976) 1075-1082.

[2] Shortliffe, E. H. and Buchanan, B. G., A Model of Inexact Reasoning in Medicine, in B G. Buchanan and E. H. Shortliffe, (Eds.), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley, Reading, MA, 1984) 233-262.

[3] Adams, J. B., Probabilistic Reasoning and Certainty Factors, in B. G. Buchanan and E. H. Shortliffe, (Eds.), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley, Reading, MA, 1984) 263-271.

[4] Gordon, J and Shortliffe, E. H., The Dempster-Shafer Theory of Evidence, in B. G. Buchanan and E. H. Shortliffe, (Eds.), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley, Reading, MA, 1984) 272-292.

[5] Barnett, J. A., Computational Methods for a Mathematical Theory of Evidence, in *Proceedings 7th International Joint Conference on Artificial Intelligence*, Vancouver, (BC 1981) 868-875.

[6] Gordon, J. and Shortliffe, E. H., A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space, *Artificial Intelligence* 26 (1985) 323-357.