

대화형 에이전트의 주제추론을 위한 스크립트 적응적 베이지안 네트워크 자동 생성

임성수⁰ 조성배

연세대학교 컴퓨터과학과

lss@scslab.yonsei.ac.kr⁰, sbcho@cs.yonsei.ac.kr

Automatic Construction of Script-adaptive Bayesian Networks for Topic-Inference of Conversational Agent

Sungsoo Lim⁰ Sung-Bae Cho

Dept. of Computer Science, Yonsei University

요 약

인터넷을 통한 정보 제공이 늘어남에 따라서 사용자가 원하는 정보를 손쉽게 얻기 위한 연구가 활발히 진행되고 있으며, 이러한 연구 중 하나가 대화형 에이전트이다. 최근 대화형 에이전트에서 사용자 질의의 주제 추론을 위하여 베이지안 네트워크가 적용되었다. 하지만 베이지안 네트워크의 설계는 많은 시간이 소요되며, 스크립트(대화를 위한 데이터베이스)의 추가변경시에는 베이지안 네트워크도 같이 수정해야 하는 번거로움이 있어 대화형 에이전트의 확장성을 저해하고 있다. 본 논문에서는 스크립트로부터 베이지안 네트워크를 자동으로 생성하여 베이지안 네트워크를 이용한 대화형 에이전트의 확장성을 높이는 방법을 제안한다. 제안하는 방법은 베이지안 네트워크의 구성 노드를 계층적으로 설계하고, Noisy-OR gate를 사용하여 베이지안 네트워크의 조건부 확률 테이블을 계산한다. 피험자 10명이 대화형 에이전트를 위한 베이지안 네트워크를 수동 설계한 것과 비교하여 제안하는 방법의 유용성을 확인하였다.

1. 서론

인터넷 사용이 보편화됨에 따라서 인터넷을 통하여 많은 정보가 쏟아져 나오고 있다. 이러한 정보의 홍수 속에서 사용자와 정보를 주고받을 수 있는 효과적인 방법의 필요성이 크게 늘고 있다. 이에 따라서 사용자가 원하는 정보를 제공해주기 위해 최근 자연어로 정보를 주고받으며 사용자에게 필요한 정보를 제공하는 대화형 에이전트에 관한 연구가 진행 중이다. 제안하는 대화형 에이전트에서는 사용자 질의 의도를 파악하기 위하여 패턴 매칭, 베이지안 네트워크 주제 추론 등의 기법을 사용한다[1,2]. 그런데 베이지안 네트워크의 설계는 초보자들이 쉽게 접근하기 어려울 뿐만 아니라 많은 시간이 소요된다. 그리고 이러한 베이지안 네트워크는 대화형 에이전트에서 스크립트의 추가-변경이 일어나는 경우 부분적인 수정이 필요하므로 매우 번거롭다.

본 논문에서는 대화형 에이전트 설계에 드는 시간 비용을 감소시키고, 초보자들도 쉽게 에이전트 설계에 접근할 수 있도록 베이지안 네트워크의 자동 생성기법을 제안하고, 실험을 통하여 제안하는 방법의 유용성을 검증한다.

2. 관련연구

2.1 대화형 에이전트

대화형 에이전트는 자연어로 해당 분야의 전문 지식을 알려주는 메신저 기반의 대화형 전문가 시스템이다. 인간이 서로 의사소통 수단으로 사용하는 자연어를 인간-컴퓨터간의 통신에 사용한다면, 기존의 컴퓨터 제약적이거나 사용자 제약적인 시스템에서 제공하는 메뉴와 같은 정보전달방식과는 달리, 상호작용을 풍부하게 하고 단순히 하나의 단어나 사용자 입력을 통해서 전달하는 정보에 비해 훨씬 복잡한 정보를 포함할 수 있다[3, 4].

최초의 대화형 에이전트인 ELIZA[5]는 사용자로부터 입력되는 문자에서 특정한 키워드와 구 집단을 단순 패턴 매칭 기술을 이용해 얻어냄으로써 응답을 수행한다. 그리고 AIML(Artificial Intelligence Markup Language)을 사용한 ALICE[6]는 순차적 패턴 매칭 기법을 사용함으로써 사용자질의에 대한 분석의 수

준을 높였다. 하지만 이러한 패턴매칭만으로는 사용자 의도 파악에 한계가 있다. 따라서 사용자 의도를 보다 잘 파악하기 위해서 오토메타를 사용한 사용자 질의문형 분석 및 베이지안 네트워크를 통한 주제추론 등의 기법 등이 적용되었다[2].

2.2 베이지안 네트워크 학습

베이지안 네트워크는 방향성 비순환 그래프(directed acyclic graph, DAG)로 그래프의 각 정점(vertex)은 확률변수를 나타내며, 아크는 정점들 간의 의존성을 나타낸다. 베이지안 네트워크는 불확실한 상황에서 정보를 표현하거나 추론하는 대표적인 방법으로 인공지능 분야에서 널리 사용되고 있다.

베이지안 네트워크를 학습하는 문제는 주어진 평가 척도에 따라 데이터의 훈련 집합(training set) D 에 가장 잘 부합되는 네트워크를 구하는 것이며, 여기서 D 는 모든(또는 적어도 몇몇) 변수에 대한 값의 사례 집합이다. 네트워크를 구한다는 것은 DAG 구조와 DAG의 각 노드에 연관된 조건부 확률 테이블(conditional probability table, CPT)을 함께 구하는 것을 의미한다[7]. 만일 네트워크의 구조가 알려져 있다면 CPT만을 구하면 된다. 하지만 CPT를 구하는 문제는 난해하여 전문가들도 문제 영역에 알맞은 구조만 구하고 CPT는 얻지 못하는 경우가 종종 있다. 만일 문제 영역에 해당하는 표본 통계 자료가 준비되어 있다면, 조건부 확률을 이용하여 CPT의 값을 구할 수 있다. 하지만 보다 현실적인 설정에서는 결여된 데이터가 존재하며 이러한 결여된 데이터는 CPT의 학습을 어렵게 만든다[7].

베이지안 네트워크의 DAG 구조가 알려져 있지 않은 경우에는 네트워크 구조를 평가하는 평가척도를 도입하여 네트워크 구조를 탐색하며 가장 적합한 네트워크를 얻어야 한다. 하지만 모든 네트워크를 탐색하는 것은 NP-hard임이 증명되었다[8]. 따라서 모든 네트워크를 탐색하는 것은 불가능하므로 유전자 알고리즘과 같은 휴리스틱한 방법을 사용한다[9].

3. 제안하는 방법

본 논문에서 개발한 대화형 에이전트의 답변 수행과정은 그

림 1과 같다.

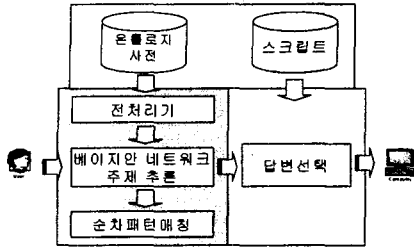


그림 1. 대화형 에이전트 구조도

전처리 단계에서는 형태소 분석을 통하여 명사와 동사의 어근을 핵심 단어로 추출한 후 동의어 사전을 사용하여 동의어로 변환하는 역할을 한다. 그리고 선택된 핵심 단어들을 기반으로 페이지안 네트워크를 통한 확률적 주제 추론을 하여 주제를 선택하고[3], 핵심어와 스크립트의 키워드간의 순차 패턴 매칭을 통하여 적절한 답변이 선택된다[1]. 표 1은 본 논문에서 사용된 대화형 에이전트 스크립트의 형식을 나타내고 있다.

표 1. 스크립트의 형식

```
[letter] := 영어 | 한글
[word] := [letter]+
[script_file] := [script]+
[script] := "<script>" [topic] [class] [keyword]+ [answer]+ "</script>"
[topic] := "<topic>" [word]+ "</topic>"
[keyword] := "<keyword>" [word]+ "</keyword>"
[answer] := "<answer>" [word]+ "</answer>"
```

3.1 페이지안 네트워크 구조 생성

제안하는 대화형 에이전트에서 사용되는 페이지안 네트워크는 비교적 인과관계가 명확하므로 그 구조를 규칙에 의해서 계층적으로 생성한다. 네트워크의 상위계층은 스크립트의 [topic]으로 이루어진 주제 노드로 페이지안 추론을 통하여 얻고자하는 확률 변수가 된다. 하위계층은 스크립트 [keyword]에서의 하나의 [word]로 이루어진 키워드 노드로 이는 증거노드가 된다. 그리고 중간계층은 소주제 노드와 중간주제 노드로 구성되며, 스크립트에서 [keyword] 하나가 하나의 소주제 노드를 구성하고, 중간주제 노드는 네트워크의 크기가 커짐에 따라서 생성된다.

표 2. 스크립트의 예

```
<SCRIPT>
<TOPIC> Agent Age </TOPIC>
<KEYWORD> 너 나이 </KEYWORD>
<KEYWORD> 너 몇살 </KEYWORD>
<ANSWER> 한 살인데요. </ANSWER>
<ANSWER> 태어난지 몇 일 안됐어요. </ANSWER>
</SCRIPT>
```

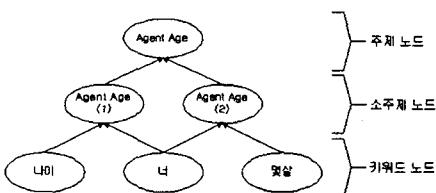


그림 2. 페이지안 네트워크 구조

그림 2는 표 2의 스크립트를 페이지안 네트워크로 구성한 것이다. 스크립트의 [keyword]는 소주제 노드가 되어 계속해서 해당 주제 노드의 하위노드로 추가된다. 페이지안 네트워크는 상위노드의 확률 값을 계산할 때 하위노드의 지수 승에 비례하는 연산량이 필요하다. 만일 하나의 주제 노드에 여러 개의 소주제 노드들이 추가되면 많은 양의 연산이 필요하다. 제안하는 방법에서는 연산량이 지수적으로 증가하는 것을 막기 위해서 하위노드의 개수를 제한한다. 즉, 하위노드를 n개 이상 갖지 못하며, 만약 하위노드의 개수가 n개가 되면, B트리와 같은 방식으로 노드를 분화한다. 그림 3은 n=4인 경우의 분화과정을 보여준다.

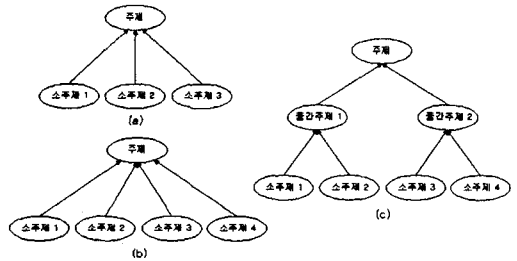


그림 3. 페이지안 네트워크 구조 생성과정
(a) 초기상태, (b) 소주제 추가, (c) 노드 분화

노드를 분화하지 않는 경우에는 소주제 노드의 개수가 k개 인 경우 주제 노드의 확률 값을 구하기 위한 연산량은 2^k이 된다. 그러나 노드를 분화시키면, 네트워크를 키워드 노드를 제외한 완전트리로 가정하면 트리의 높이가 h이고 각 노드의 하위노드의 개수는 n-1일 때, k=(n-1)^h이 된다. 따라서 확률 값 계산이 필요한 주제노드와 중간주제 노드들의 개수는 (1)이 되고, 각 노드의 하위노드 개수가 n-1개가 되므로, 총 연산량은 (2)가 된다. 즉, 노드의 분화를 통해서 연산량이 O(2^k)에서 O(k)로 바뀐다.

$$\frac{(n-1)^h - 1}{n-2} = \frac{k-1}{n-2} \tag{1}$$

$$\frac{k-1}{n-2} \times 2^{n-1} \tag{2}$$

3.2 조건부 확률 테이블(CPT) 생성

대화형 에이전트에서는 대화에 관한 통계 자료가 준비되어 있지 않으므로 CPT를 생성하기 위하여 페이지안 네트워크에서 널리 사용되고 있는 Noisy-OR gate를 사용한다[10]. Noisy-OR gate에서는 x_i:원인(하위노드), y:결과(상위노드)라고 하고, p₀는 x가 모두 나타나지 않았을 때 저절로 y가 나타날 확률, 그리고 p_i는 x_i만 나타나고 나머지 원인은 나타나지 않았을 때 결과가 나타날 확률이라고 하면, X_p가 나타났을 때 결과가 나타날 확률은 다음과 같이 정의된다.

$$Pr(y | X_p) = 1 - (1 - p_0) \prod_{i: x_i \in X_p} \frac{1 - p_i}{1 - p_0}$$

여기서는 p₀의 값으로 0.001을 주었으며, 주제노드와 중간노드에서의 확률 변수 값은 다음과 같이 정의한다. n은 하위노드의 개수이고 a는 가중치를 의미한다.

$$p_i = a + \frac{1-a}{n}$$

그리고 소주제 노드에서의 확률 변수 값은 다음과 같이 정의한다. m은 하위노드 x_i의 상위노드의 개수이다.

$$p_i = \frac{1-a}{m}$$

위의 식에서 보면 n 으로 나누어 주는 부분이 있는데, 이는 각 하위노드가 상위노드에 미치는 영향이 같다는 가정 하에서 각 하위노드가 상위노드에 미치는 영향을 고르게 분배하는 역할을 한다. 또한 소주제 노드의 경우에는 n 뿐만아니라 m 으로도 나누어주는데, 이는 키워드가 여러 개의 상위노드를 갖는다면 다른 노드에도 함께 영향을 주므로 상대적으로 그 비중을 줄여줄 필요가 있기 때문이다. 그리고 a 값은 상위노드가 주제 노드에 가까울수록 하위노드들이 주제노드에 미치는 영향이 커지므로 높고, 그렇지 않을수록 낮은 값을 주어진다. 본 논문에서는 a 값을 소주제 노드에서는 0.5, 주제 노드에서는 0.999로 설정하였다.

4. 실험 및 결과

대화형 에이전트에서의 베이지안 네트워크는 사용자 질의에 해당하는 주제 노드에서의 확률 값(T_0)이 클수록, 그리고 T_0 와 T_k (다른 주제 노드에서의 확률 값)와는 차이가 많이 날수록 좋은 성능을 갖는다. 따라서 본 논문에서는 제한한 방법의 성능 평가를 위해서 표 3과 같은 평가 척도를 사용하였다.

표 3. 적합도 평가 방법

T_0 := 사용자 질의가 의도한 주제의 확률 값
T_k := 각 주제의 확률 값 ($1 \leq k \leq N$)
N := 주제의 개수 - 1
적합도 $M = (T_0 - \sum_{k=1}^N D_k) \times 100$ (%)
$D_k = \begin{cases} \text{if } (T_0 < T_k) & \frac{1}{N} \\ \text{else if } (T_0 - T_k < 0.1) & \frac{0.1 - (T_0 - T_k)}{0.1 \times N} \\ \text{otherwise} & 0 \end{cases}$

실험은 컴퓨터 전공의 대학생 10명을 대상으로 학생들이 직접 설계한 베이지안 네트워크와 제한한 방법으로 설계한 베이지안 네트워크를 비교하였다. 실험에 사용된 스크립트는 베이지안 네트워크 설계에 많은 시간이 소모된다는 점을 감안하여 11개의 주제를 갖는 스크립트만을 사용하였다. 그림 4는 실험결과를 나타내고 있다.

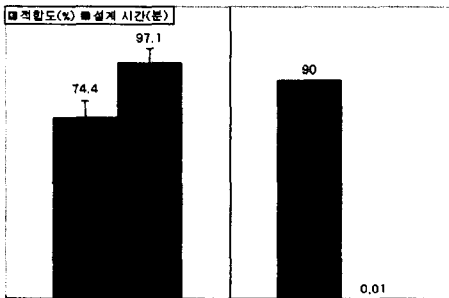


그림 4. 수동 설계와 자동 설계 비교

실험 결과 피험자들은 평균 74.4%의 적합도를 가지는 베이지안 네트워크를 평균 97.1분에 설계하는 것으로 나타났다. 반면에 제안하는 방법은 90%의 적합도를 나타냈으며 설계시간은 수 초에 불과하였다.

5. 결론

베이지안 네트워크를 사용하는 대화형 에이전트의 설계에 있어서 베이지안 네트워크의 설계는 필수적이다. 그러나 베이지안 네트워크는 초보자들이 쉽게 접근하기 어려우며, 전문가들도 네트워크 설계에 많은 시간이 필요하다.

본 논문에서는 대화형 에이전트에서의 베이지안 네트워크를 자동 생성하는 방법을 제안하고 그 성능을 수동 설계와 비교 평가해본 결과 초보자에게 비하여 더 좋은 성능의 베이지안 네트워크를 짧은 시간에 자동으로 생성함을 확인하였다. 이러한 방법을 사용하면 대화형 에이전트의 설계에 드는 시간과 비용을 크게 줄이며 초보자도 대화형 에이전트에 쉽게 접근할 수 있을 것이다.

향후에는 베이지안 네트워크의 자동 설계뿐만 아니라 베이지안 네트워크에 학습 기능을 넣어서 보다 정확한 주제 추론을 가능하도록 하는 기능이 요구된다. 그리고 초보자들이 손쉽게 대화형 에이전트를 설계할 수 있도록 하는 보다 편리한 인터페이스가 요구된다.

감사의 글

이 연구는 과학기술부가 지원한 뇌과학 연구 프로그램에 의해 지원되었음.

참고문헌

- [1] 김수영, 조성배, "순차적 패턴매칭 기법을 이용한 대화형 도우미 에이전트," *한국정보과학회 2000 추계학술발표*, vol. 27, no. 2, pp. 24-26, 2000.
- [2] 홍진혁, 조성배, "계층적 베이지안 네트워크를 이용한 대화형 에이전트의 문맥유지," *추계 정보과학회 학회지*, vol 29, no. 2, pp. 259-261, 2002.
- [3] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu and A. Stent, "Towards conversational human-computer interaction," *AI Magazine*, vol. 22, no. 4, pp. 27-38, 2001.
- [4] P. Nugues, C. Godereaux, P.O. El-Guedj, and F. Revolta, "A conversational agent to navigate in virtual worlds," *Proc. 11th TWLT, Dialogue Management in Natural Language Systems*, pp. 23-33, 1996.
- [5] J. Weizenbaum, "ELIZA - A computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36-45, 1965.
- [6] The A.L.I.C.E. AI Foundation, In <http://alicebot.org>
- [7] D. Heckerman, "A tutorial on learning with Bayesian networks," *Microsoft Research, Technical Report MSR-TR-95-06*, 1995
- [8] D. Chickering, D. Geiger and D. Heckerman, "Learning Bayesian networks is NP-hard," *Proc. 5th Conf. on Artificial Intelligence and Statistics*, pp. 112-128, 1995.
- [9] R. Etxeberria, P. Larranaga and J.M. Picaza, "Analysis of the behaviour of genetic algorithms when learning Bayesian network structure from data," *Pattern Recognition Letters*, vol. 18, Issues 11-13, pp 1269-1273, November 1997.
- [10] A. Onisko, M. J. Druzdzal and H. Wasyluk, "Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates," *Int. Journal of Approximate Reasoning*, vol. 27, Issue 2, pp 165-182, August 2001.