

자기 구성 지도를 이용한 전자메일 자동응답 및 브라우징 시스템의 사용성 평가

노영주⁰ 조성배
연세대학교 컴퓨터과학과
(yjnoh, sbcho)@candy.yonsei.ac.kr

A Usability Test of E-mail Automatic Response and Browsing System Using Self-Organizing Map

Young-Ju Noh⁰ and Sung-Bae Cho
Dept. of Computer Science, Yonsei University

요 약

컴퓨터의 사용인구가 많아지고, 인터넷의 보급이 급속히 늘어남에 따라 많은 정보가 생산되고 있다. 그리고 이러한 정보를 사용자에게 좀더 효율적인 방법으로 제공하는 서비스들도 많아지게 되었다. 그러나 컴퓨터에 익숙하지 않은 사용자들은 쉽게 이러한 서비스를 이용하지 못하기 때문에 사용자를 돕는 시스템이 필요하다. 인터넷 서비스제공 업체들은 사용자의 질문에 대해 관리자가 직접 답을 해주는데, 이들 시스템을 이용하는 사용자들의 증가로 질의응답 업무의 양이 커지고 있다. 본 논문에서는 이를 해결하기 위해 사용자의 질의를 자동으로 분류하여 응답하고 사용자가 FAQ를 개념적으로 브라우징 할 수 있는 시스템의 유용성을 입증하기 위하여, 그 적용 가능성과 일반 사용자들의 이용 결과를 통계적으로 분석하였다.

1. 서 론

컴퓨터와 PC통신의 폭넓은 보급과 인터넷에 대한 관심의 확산이 많은 사람들을 정보통신 기반 서비스로 끌어들이고 있다. 그러나, 이러한 서비스에 익숙하지 않은 사용자가 자신이 원하는 정보를 찾는 것은 쉬운 일이 아니다. ISP, PC통신업체 등 정보통신 서비스 업체는 이러한 문제를 해결하기 위해서 전화상담 창구를 운영하고, FAQ나 게시판의 형태로 유형화된 질문에 대한 답을 제공하기도 하며, 전자우편으로 사용자의 질문에 대한 응답을 준다. 그러나 사용자들의 증가로 인해 답을 제공하는 서비스는 많은 인력을 필요로 하게 되었고, 질의응답의 자동화의 필요성이 대두되었다.

한메일넷의 경우, 사용자는 2001년 현재 1천만명 이상의 사용자가 이용하고 있으며, 하루 평균 200 통 정도의 사용자 질의를 처리하고 있는데, 이를 실시간으로 자동 응답한다면 사용자에게 만족도 높은 서비스를 제공 할 수 있을 것이다. 뿐만 아니라 관리자도 중복된 일을 피할 수 있으므로 효율적인 일 처리가 가능하다. 따라서, 사용자와 관리자의 편의를 도모하기 위해서 질의 자동 응답 시스템은 개발할 필요가 있다.

본 논문에서는 SOM(Self Organizing Map)을 이용하여 사용자 질의 메일을 자동으로 응답하는 시스템과 사용자가 직접 답변을 검색할 수 있는 브라우징 시스템도 제안하였다. 이러한 시스템을 통해 전자 우편의 자동 응답의 적용 가능성과 브라우징의 효율성을 비교하고자 한다.

2. 브라우징 시스템

시스템의 기능은 크게 두 가지다. 자동응답의 경우, SOM(Self Organizing Map)을 통해 질의 메일을 분류하여 적절한 답변 메일을 전송하는 작업을 한다. 그리고 브라우징의 경우, 학습된 SOM을 키워드들로 표시하고 다중 레벨로 검색할 수 있도록 하여 사용자가 자신이 찾고자하는 답변을 쉽게 검색할 수 있게 한다. 이 두 시스템의 기본이 되는 신경망은 SOM인데, 데이터 축약을 위해서 메일에 대한 전처리

과정과 각 메일을 분류하는 시스템을 구축할 때 사용이 된다.

2.1 키워드 추출

질의 메일들은 자연 언어로 이루어져 있다. 질의 메일을 신경망의 입력으로 인코딩하기 위해서는 정규화된 벡터의 형태로 변화시켜야 한다. 벡터화를 위해서 먼저 질의에서 의미 있는 키워드만을 추출하여 불용어들과 불필요하게 반복되는 키워드들을 제거하게 된다.

2.2 데이터 축약

키워드의 집합을 수치화 된 벡터로 표현하는 작업이 필요하다. 기본적으로 벡터 스페이스 모델을 포함한 여러 가지 방법들이 있으나 질의 메일들은 키워드의 수가 아주 많기 때문에 기존의 방법들을 통한 인코딩은 적절하지 못하다. 그래서 SOM(Self Organizing Map)을 이용한 인코딩 방법을 사용하였다. 여기서 사용되는 SOM은 쉽게 말해서 동의어 사전으로서의 역할을 한다. SOM의 입력값은 각 단어들에 대한 문맥 정보들이 되고 결과는 문맥 정보에 의해 분류된 키워드들이 된다. 이 경우, SOM의 특징에 의해서 유사한 키워드들은 SOM의 같은 노드에 할당되거나 근접한 위치에 할당되게 된다[1]. 이렇게 완성된 데이터 축약 자기 구성 지도는 2.3에서의 입력 벡터를 생성하는데 이용된다.

2.2.1 문맥 정보 인코딩

SOM의 입력으로 문맥 정보가 이용되는데, 이 문맥 인코딩하는 방법은 다음의 수식으로 표현 될 수 있다[4].

$$X(i) = \begin{bmatrix} E\{x_i - 1 | x_i\} \\ \varepsilon x_i \\ E\{x_i + 1 | x_i\} \end{bmatrix} \quad (1)$$

(1)에서 키워드 x_i 에 대한 입력 벡터 $X(i)$ 는 x_i 의 선행자와 후행자의

평균으로 구성되어 있다. 여기서, 선행자의 평균은 질의 자료의 모든 x_i 에 대하여 x_i 의 바로 앞에 나오는 키워드들의 벡터 값을 합하여 평균을 구한 값이다. 그리고 후행자는 자료의 모든 x_i 의 바로 뒤에 나오는 키워드들의 벡터 값을 합하여 평균을 구한 값이다. 선행자와 후행자는 x_i 에 대한 특징을 나타내는 값이다. 모든 데이터에 대해서 x_i 의 앞과 뒤에 나오는 키워드들을 살펴봄으로써, 문맥 정보를 얻을 수 있다[2]. 이 방법은 기본적으로 하나의 키워드의 앞과 뒤에 나타날 수 있는 키워드들은 그 키워드의 문법적인 특징과 각 질의 메일의 내용에 따라 항상 거의 일정하게 나타난다고 가정에 의해 도출되었다. 이 입력벡터를 SOM에 입력하게 되면 결과로 문맥 정보에 의해 분류된 자기 구성 지도가 산출된다.

2.3 입력 벡터 생성

입력 벡터는 데이터 축약 자기 구성 지도의 결과를 통해서 생성된다. 일단 데이터 축약 자기 구성 지도가 만들어 진 다음에는 자기 구성 지도의 각 노드에 어떤 키워드들이 매핑되어 있는 지를 알 수 있다. 이 경우, 각각의 질의 메일의 키워드들이 데이터 축약 자기 구성 지도의 몇 번째에 할당되는 지를 알 수 있게 되고, 각 키워드들에 대한 히스토그램을 구할 수 있게 된다. 예를 들면, 데이터 축약 자기 구성 지도가 이라고 하고, 질의 메일의 키워드들이 각각, (0,0)에 2번 (1,2)에 1번 나타났다고 한다면, 인코딩은 다음과 같이 된다.

2	0	0	0	0	1	0	0	0
---	---	---	---	---	---	---	---	---

앞에서와 같이 인코딩 된 입력 벡터의 차원은 자기 구성 지도의 크기와 같게 된다. 즉, $m \times n$ 의 자기 구성 지도에 의해 생성될 수 있는 입력 벡터는 mn 의 1차원 벡터가 되는 것이다. 여기서 만들어진 각각의 히스토그램은 각 질의 메일 하나 하나의 고유의 입력 벡터가 된다.

2.4 질의 응답 시스템

2.3에서 만들어진 입력 벡터를 문서 분류 SOM으로 학습하게 되면, 각 질의 메일은 SOM의 특정 노드에 매핑이 되게 된다. SOM의 특징에 의하여, 같은 클래스의 질의들은 같은 노드나 근접한 위치에 있는 노드에 매핑 될 것이다. 이렇게 학습된 SOM에 새로이 사용자가 보낸 질의 메일을 인코딩하여 입력하게 되면 그 질의 메일이 어떤 노드에 매핑되는 지를 알 수 있게되고, 그 노드가 나타내는 클래스에 대한 답변 메일을 전송하게 되는 것이다.

2.5 브라우징 시스템

브라우징 시스템은 학습된 자기 구성 지도를 시각화(visualization)함으로써 사용자가 자신이 알고 싶은 질의에 대한 답변을 직접 검색할 수 있도록 해준다. 브라우징 시스템은 3개의 레벨을 가진다. 첫번째 레벨은 자기 구성 지도 전체를 보여 준다. 여기에서 지도는 부분 적으로 각 노드의 클래스를 대표할 수 있는 키워드들로 표시되어 있다. 두번째 레벨은 첫 번째 레벨에서 클릭된 부분을 지역적으로(locally) 확대하여 세부적으로 검색할 수 있도록 해주고, 세번째 레벨도 마찬가지로 역할을 하게 된다. 그리고 세번째 레벨에서 특정 노드를 클릭하게 되면 그 노드의 클래스에 대한 답변을 출력한다.

키워드 표시는 각 노드의 클래스를 가장 잘 대표할 수 있는 키워드를

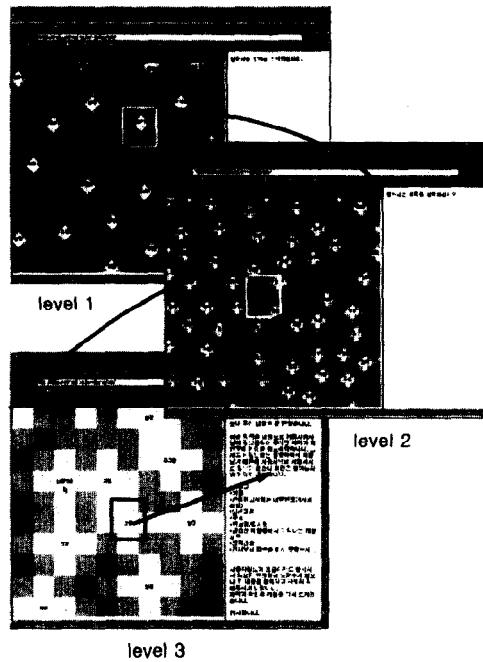


그림 1. 단계별 브라우징 시스템

산출하여 나타내어야 한다. 기본적으로 각 노드에 매핑 된 질의 문서들의 키워드의 빈도수를 통해 대표 키워드를 계산한다. 그러나 어떤 키워드는 단 한번 나타나더라도 하나의 클래스에 대한 특징을 나타낼 수 있는데 비해, 높은 빈도수를 가지더라도 클래스에 대한 특징을 나타내지 못하는 키워드들도 있다. 그래서 (2)를 통해서 빈도수에 대한 대표치를 보완하였다.

$$I_w = F \times \frac{DocFin}{DocFout} \quad (2)$$

여기서, I_w 는 키워드 w 의 중요도를 나타내는 값인데, F 는 노드내의 w 의 빈도수를 나타내고, $DocFin$ 은 노드내에 w 를 가지는 문서의 수, $DocFout$ 은 노드밖에서 w 를 가지는 문서의 수를 각각 나타낸다.

3. 브라우징 시스템의 사용성 평가

시스템의 사용성 평가는 설문을 통하여 이루어졌다. 제안된 브라우징 시스템은 IBM사가 제안하는 사용자 중심의 편리한 사용을 위한 디자인 방법[5]을 적용하였다.

- 사용자의 요구사항이 잘 반영되었는지 파악한다.
- 인터페이스에 사용자가 능숙하게 사용하는 보편적인 인터페이스를 활용한다.
- 프로그래밍의 환경(기술, 프로그래머 능력 등등)이 가능한 정도를 활용한다.

표 1 사용자 중심의 디자인방법

제시된 시스템의 사용성을 평가하기 위해 표 2 와 같은 보편화된

기준을 설정하였다.[6]

▷ 사용자 이용성의 보편화
시스템이나 프로그램에 구애를 받지 않고 일반적인 모든 사용자가 같은 방법으로 이용할 수 있게 만들어야 한다.
▷ 융통성 있는 브라우징 디자인
사용자 개개인에 대한 능력의 차이나 선호도를 포용한다.
▷ 간단하고 직관적인 사용성
사용자가 경험이나 학력, 언어능력 또는 집중력 정도와 상관없이 쉽게 이해하여야 한다.
▷ 정보의 명확한 전달
원하는 정보를 정확히 모든 사용자에게 알려야 한다.
▷ 사용자에게 세심한 배려
사용자의 실수에 대한 경고 및 복구요소를 제공하여 불필요하게 수반되는 불안 요소를 제거하고 신뢰성을 확보한다.

표 2. 평가기준

조사는 기준을 바탕으로 만들어진 설문지를 통해 전자메일을 사용하는 사람들을 대상으로 조사하였다. 조사 방법은 사용자가 제안된 시스템을 통해 찾고자 하는 의문점을 메일로 보내서 질의에 맞는 답신을 받거나 질문에 대한 답을 직접 브라우징으로 찾아가는 방식을 택했다. 조사 대상의 특성은 표 3 과 같다.

변수	변수값	빈도수	%
성별	남자	177	70.23
	여자	85	29.77
연령별	10대	74	29.36
	20대	127	50.39
	30대	38	15.07
	40대 이상	13	5.18
학력수준	초등학생	13	5.18
	중학생	32	12.69
	고등학생	29	11.50
	대학생	141	55.95
	대학원 이상	37	14.68
인터넷 사용기간	6개월 미만	9	3.57
	6개월~1년	8	3.17
	1년~2년	37	14.68
	2년 이상	198	78.58
총 계		252	100

표 3. 응답자 특성

비교 대상으로 정한 것은 같은 텍스트를 이용하여 기존의 디렉터리 형식으로 질문의 답을 찾아가는 방식을 택하였다. 이때 각각의 만족

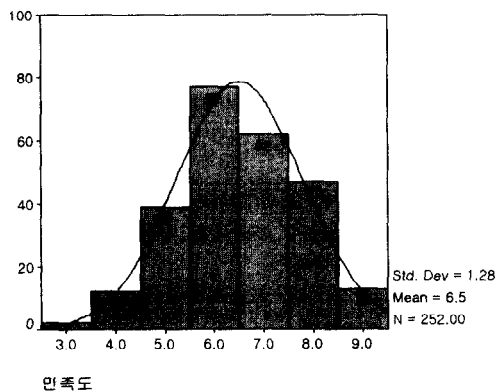


그림 3. 계층적 브라우징에 대한 사용자 만족도 분포

도를 조사하여 연령, 성, 학력수준, 인터넷 사용기간별 분포를 조사하였다.

조사 결과 계층적 브라우징의 전체적인 만족도를 나타내는 그림3의 그래프를 분석해 보면 응답자들은 평균 65%의 만족도를 보이고 있다. 조사된 각각의 변수에 따라 대부분의 사람들은 제안된 시스템에 대해 50~80%사이의 만족도를 보인다. 이와같은 수치는 제안된 브라우징 시스템이 흥미롭게 접근할 수 있고, 사용자에게 원하는 답변을 찾을 수 있다는 가능성을 제시해준다.

또한 기존 디렉터리 방식의 시스템과 달리 자바 애플릿을 사용해 계층적인 인터페이스를 구현함으로써 초기 시스템 구동 시간은 좀 많이 걸리지만 검색 시에는 서버와의 접속 없이 검색을 할 수 있게 되어서 여러 번 검색을 할 경우에 좋은 성능을 보였다.

4. 결 론

본 연구에서는 SOM을 이용하여 한메일넷 질의 메일들에 대한 자동응답 시스템과 질의 브라우징 시스템에 대해서 살펴보았다.[7] 학습된 SOM을 이용한 브라우징 시스템의 경우, 기존 단순한 키워드 나열을 통한 검색이 아니라 사용자가 문서의 전체적인 분포를 파악하고, 손쉽게 자신의 질의를 검색할 수 있게되었다. 또 검색된 결과 역시 학습 데이터의 분류율 만큼의 정확도를 가지고 있으므로 효율적인 시스템이라고 할 수 있다. 특히 사용자들에게 직접 사용을 한 후 사용성을 평가한 것에서는 각 사용자들이 대체적으로 Q/A를 찾아가는데 흥미를 충분히 유발하면서 질문에 대한 답을 알 수 있었다.

참고문헌

- [1] G. Salton, *Automatic text processing : the transformation, analysis, and retrieval of information by computer.* Addison-Wesley, 1988.
- [2] H. Ritter, T. Kohonen, "Self-organizing semantic maps," *Biol Cyb*, 1989, 61:241-254
- [3] T. Kohonen, *Self-Organizing Maps*, Springer, Berlin Heidelberg, 1995.
- [4] S. Kaski, T. Honkela, Krista Lagus, T. Kohonen, "Creating an Order in Digital Libraries with Self-Organizing Maps," *WCNN'96*, pp. 814-817
- [5] IBM, "Web Design Guide Line", http://www-3.ibm.com/ibm/easy/eou_ext.nsf/publish/572, 2001.
- [6] 송승현, 기우서, 김용곤, "사용자 중심의 웹 인터페이스 디자인 및 평가에 관한 연구", *한국정보과학회*, 2001. 4, pp. 478~480
- [7] 김현돈, 조성배, "한메일넷 질의 자동응답을 위한 이단계 자기 구성지도", *한국정보과학회*, 2000. 4, pp. 481~483