

유비쿼터스 환경에서 키워드링크방식을 이용한 관련서비스 제공 기술

최재홍[○], 옥지웅, 김응모
성균관대학교 컴퓨터공학과

zoozung[○]@nate.com, {okjwguy, umkim}@ece.skku.ac.kr

Related Service Offers Technology with Keyword Link Method in Ubiquitous Environment

Jaehong Choi[○], Jeewoong Ok, Ungmo Kim
Dept. of Computer Engineering, SungKyunKwan University

요 약

유비쿼터스 컴퓨팅 환경에서 사용자가 원하는 데이터를 제공하는 일은 중요한 일이다. 또한, 급격히 늘어나는 정보를 효율적으로 저장하는 것도 중요하다. 그러나 기존의 데이터 제공방식은 사용자의 의도를 파악하지 못한 단순 1차 정보 및 서비스를 제공하고 있다. 그리고 제공되는 정보의 신뢰도는 다수의 사용자가 사용하기엔 부족한 편이다. 따라서 본 논문에서는 더 효율적으로 데이터와 원하는 서비스를 제공하기 위해 키워드링크방식을 이용한 관련서비스 제공 기술을 연구한다. 세부적인 연구내용은 첫째, 데이터를 저장 시에 키워드를 하나씩 3가지로 나누어 가중키워드를 선정하여 part2에 저장시키고 상대적으로 가중치가 낮은 두 개의 키워드는 각각 part1, part3에 저장하여 가중키워드와 가중치가 낮은 키워드를 서로 링크로 연결시키고 둘째, 데이터 마이닝을 통한 정보 및 서비스를 제공할 때 검색한 데이터 외에 키워드링크방식을 통하여 관련된 데이터를 2개 이상 제공하여 다수의 사용자가 원하는 정보 및 서비스를 제공한다. 따라서 본 논문에서는 사용자의 정보를 좀 더 효과적으로 저장하고, 데이터마이닝할 수 있는 방법을 제안하고 있다.

1. 서 론

유비쿼터스 개념이 우리 생활에 점차 도래함에 따라 이러한 환경이 필요로 하는 데이터의 수집과 배포의 요구가 점점 커지고 있다. 지금까지는 사용자가 어떤 정보를 얻거나 서비스를 이용하기 위해서는 그 정보나 서비스에 대해 적당한 지식이 필요했고, 그 지식을 바탕으로 사용가능한 서비스만 이용할 수 있었다. 그러나 서비스의 양이 방대해지고, 종류가 급격히 늘어남에 따라 모든 서비스에 대해 충분히 알고 사용하기 어려운 점이 있다. 또한, 사용자들은 자신의 부분 정보만으로도 많은 정보 및 서비스를 이용하고자 하는 경우가 있으므로, 어떠한 상황에서도 사용자에게 최적의 서비스를 제공하는 것이 중요한 일로 대두되고 있다. 만약 사용자의 무지로 원치 않는 서비스를 요청할 경우 올바른 서비스를 찾을 수 있도록 관련된 다른 서비스를 제공해야 한다. 또는 원하는 서비스와 더불어 이용하면 도움이 되는 서비스가 있을 경우 시스템에서 추가적으로 관련된 데이터 및 서비스를 제공하는 것이 필요하다.

본 논문에서는 사용자에게 부분정보만으로도 효율적으

로 정보를 제공하며 서비스를 이용할 수 있는 방법을 제안한다.

2. 관련연구

이 장에서는 3장에서 다루게 될 데이터마이닝의 의미와 기법을 소개하고 연구한다.

2.1 데이터마이닝

데이터마이닝이란 자동화되고 지능을 갖춘 데이터베이스 분석기법으로 90년대 초반부터 지식발견, 정보발견, 정보수확 등의 이름으로도 소개되어 왔는데 일반적으로 "대량의 데이터로부터 새롭고 의미있는 정보를 추출하여 의사결정에 활용하는 작업"이라 정의한다[1]. 데이터마이닝은 '데이터베이스 내에서의 지식발견(Knowledge Discovery)'과 유사어로 언급되기도 하지만, 지식발견은 데이터로부터 유용한 정보를 발견하는 프로세스의 전 과정이며, 데이터마이닝은 지식발견 중에서도 데이터로부터 정보를 추출하기 위해서 기법을 적용하는 특정단계라고 정의할 수 있다[1].

2.2 데이터마이닝 기법

데이터마이닝 기법에는 연관규칙탐사(association rule mining), 연속규칙탐사(sequence rule mining), 분류(classification), 군집화(clustering) 등이 있다. 이들 데이터마이닝 기법은 특정 업종에만 국한된 것이 아니라 비즈니스의 환경과 목표, 사용 가능한 데이터의 속성에 따라 적용될 수 있는 적합한 기법들이 다르다[2]. 각 기법을 간략히 설명하면 다음과 같다.

2.2.1 연관규칙

많은 데이터마이닝 기법 중 가장 활발한 연구가 이루어지는 분야로서, 데이터 안에 존재하는 항목간의 종속관계를 찾아내는 작업으로, 마케팅에서는 손님의 장바구니에 들어있는 품목간의 관계를 알아본다는 의미에서 장바구니 분석이라고 한다.

2.2.2 연속규칙

연속규칙은 연관규칙에 시간관련 정보가 포함된 형태로서, 시간상에 순차적으로 나타나는 사건이나 거래의 종속관계를 찾아내는 작업이다.

2.2.3 분류

데이터마이닝에서 가장 많이 사용되는 작업으로, 부류값이 포함된 과거의 데이터로부터 부류별 특성을 찾아내어 분류모형 수립, 이를 토대로 새로운 레코드의 부류값을 예측하는 것이다.

2.2.4 데이터군집화

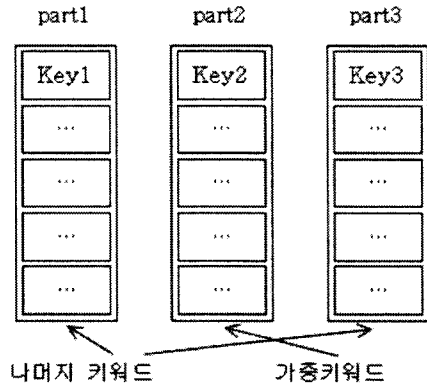
레코드를 유사한 특성을 지닌 몇 개의 소그룹으로 분할하는 작업으로서 분류작업과 흡사하나, 분석하고자 하는 데이터에 부류가 포함되어 있지 않다는 점에 차이가 있다.

3. 키워드링크를 통한 데이터 및 서비스 제공 기법

이 장에서는 효율적인 데이터 저장을 위하여 데이터를 분석하고 가중데이터를 선정하여 상대적으로 가중치가 낮은 데이터를 링크형식으로 연결하는 방법을 제안한다. 이 방법은 기존의 자료검색의 정확도를 높이는데 기여할 것으로 기대된다.

3.1 저장방법

이 서비스 제공 기법을 사용하기에 앞서 데이터베이스와 구분된 3개의 part를 생성한다. 데이터베이스에 있는 임의의 자료를 3개의 키워드로 나누어 추출하여 3개의 키워드의 중요도를 각각 비교하게 된다. 가중키워드의 선정방법으로는 데이터베이스에서 각각 3개의 키워드를 검색하여 가장 빈번히 출현하는 키워드를 가중키워드로 선정하게 된다. [그림 1]과 같이 이 가중키워드는 part2



[그림 1] 가중키워드 정렬

에 저장되고 나머지 키워드는 각각 part1, part3에 저장된다. 그 다음 각각의 part1, 2, 3에 저장된 키워드를 링크로 연결시킨다. 이는 가중키워드(Key2)를 검색하게 되면, 이와 연결된 2차 관련된 키워드를 제공하기 위함이다.

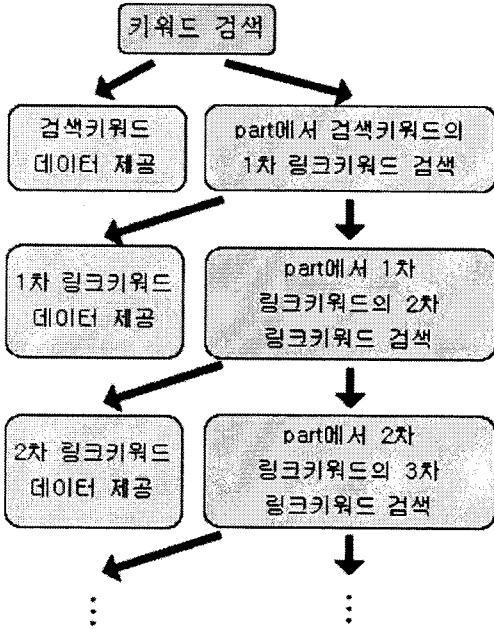
이와 같은 방법으로 데이터베이스의 모든 데이터를 키워드로 나누어 정리하여 키워드끼리의 링크를 만들 수 있다. 또한, 이후에 들어오는 데이터 역시 이와 같은 방법으로 part에 정렬시킬 수 있다. 그리고 이미 어떤 part2에 입력되어 있는 가중키워드는 다시 part2에 중복으로 입력되지 않으며, 가중키워드와 또 다른 키워드를 링크로 연결함으로써 추가적으로 관련된 키워드를 제공할 수 있게 돕는다.

하지만 같은 키워드일지라도 그 키워드가 다른 데이터에서 중요도가 낮다면 part1 또는 part3에 저장될 수 있다. 만약 이렇게 따로 키워드를 두지 않고 part 한 곳에만 그 키워드를 두게 된다면, 그물망처럼 서로 얽히게 되어서 루프(loop)가 생기게 되기 때문에 가중키워드의 역할을 못할 뿐만 아니라 정보가 정렬되지 않고 무분별하게 쌓이는 것을 방지하기 위함이다.

3.2 실행과정

[그림 2]와 같이 사용자가 데이터베이스에서 키워드를 이용하여 검색을 하면, 검색키워드에 대한 기본 검색결과를 제공하고 추가적으로 그 키워드를 이용하여 part2에 존재하는 가중키워드에서 링크된 키워드를 찾게 된다. 이 링크를 사용하여 2차 관련 키워드를 검색할 수 있고 2차 관련 키워드의 링크를 다시 찾게 된다. 그리고 만약 2차 관련 키워드에 원하는 정보가 없을 경우 추가적으로 3차 관련 키워드를 다시 2차 키워드의 링크를 이용하여 검색할 수 있다.

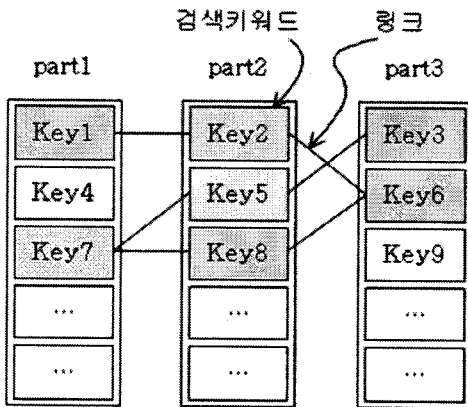
[그림 3]은 검색키워드를 이용한 추가적인 키워드 검색이 가능함을 보여준다. 검색된 Key2의 링크를 이용하여 2차 관련 키워드(Key1과 Key6)를 제공한다. 만약 이 키워드도 사용자가 원하는 것이 아니었을 경우 3차 관련



[그림 2] 키워드검색과 데이터제공 과정

키워드(Key8)를 제공하게 된다. 이와 같은 방법으로 링크가 계속 존재한다면 이 모든 관련된 키워드를 이용해서 데이터를 제공할 수 있다.

이것은 검색키워드에 국한되어 있지 않고 2차, 3차, 4차 이상의 관련 키워드를 계속해서 제공할 수 있으므로 사용자가 원하는 결과를 찾기 쉽게 도와준다.



[그림 3] 링크로 연결된 2차, 3차, 4차 관련키워드 제공

3.3 실행 평가

기존의 서비스는 키워드로 정보검색 및 서비스를 이용하고자할 때, 많은 검색결과로 인해 사용자가 원하는 정

보와 서비스를 찾기 어려웠다. 하지만 여기서 소개한 검색키워드와 관련키워드를 함께 제공한 서비스는 검색 시 가중키워드가 저장되어 있는 part2에서 그 결과를 찾아서 좀 더 사용자가 원하는 서비스에 접근했다고 할 수 있다. 예를 들면, 데이터마이닝이라는 키워드로 검색할 경우 기존의 서비스에서는 데이터마이닝을 포함하는 모든 서비스를 제공한 반면, 이 시스템에서는 데이터마이닝이 가중키워드로 선정된 정보만 제공하게 되는 것이다. 이 시스템은 핵심정보만 사용자에게 제공하기 때문에, 시간이 지날수록 데이터베이스에 점점 자료가 많이 쌓이게 되면 더 큰 효과를 볼 수 있을 것으로 기대된다. 그리고 만약 사용자의 무지로 잘못 검색해서 찾고자하는 데이터가 검색결과에 없을 경우 기존의 서비스에서는 아무런 도움을 줄 수 없었으나 이 시스템에서는 가중키워드와 링크로 연결된 2차, 3차, 4차 관련키워드를 이용하여 추가적으로 관련데이터를 제공하게 된다.

시간적으로 이 시스템의 성능을 평가하자면 기존의 서비스는 방대한 데이터 검색결과로 사용자가 찾는 데 오래 걸린 반면, 이 시스템에서는 가중키워드를 사용하여 검색하여 핵심적인 데이터와 서비스를 제공하므로 시간이 단축된다고 말할 수 있다.

정확도면에서 기존의 시스템은 검색한 키워드에만 정보를 제공하고 이 시스템은 검색한 키워드와 관련키워드를 같이 제공하기 때문에 사용자가 좀 더 원하는 결과를 제공하게 된다.

4. 결론

유비쿼터스 환경에서 다양한 사용자가 원하는 서비스를 다양한 상황에 맞게 적합한 서비스를 검색하고 제공하는 것은 중요한 일이다. 그러나 기존의 연구는 충분한 지식을 갖춘 상대의 서비스 검색만 지원하고 있고, 부분 지식으로는 적절한 서비스를 검색할 수 없고 잘못된 검색 또한 바로잡아 줄 수 없다. 유비쿼터스의 개념이 우리에게 깊이 자리잡혀가는 이 상황에서 남녀노소 상관없이 다양한 사용자가 서비스를 이용하려면 잘못된 것을 바로잡아주거나 관련서비스검색 및 제공이 절실히 필요하다.

따라서 본 논문에서는 검색키워드에 국한되지 않고 추가적으로 링크로 연결된 관련데이터를 제공함으로써, 중요하고 필요한 정보를 사용자에게 제공하기도 하고 키워드에 관련된 추가적인 서비스도 함께 제공하는 방법을 연구하였다. 그리고 검색데이터를 제공하는 시간을 줄이고 정확도도 높일 수 있음을 확인하였다. 나아가 이 연구는 효과적으로 대량의 정보를 관리할 수 있도록 도와주며, 관련된 데이터를 통합하여 새로운 데이터를 생산할 수 있게 도와줄 것이다.

향후 키워드를 3개에 국한하지 않고 트리(tree)형식으로 계속 뻗어나갈 수 있는 데이터 정렬구조를 연구하고, 사용자의 의도를 파악할 수 있도록 정확도를 좀 더 높이

는 방법을 연구하겠다.

참고문헌

- [1] 장남석외 2명, 데이터마이닝, 대청미디어, 1999.
- [2] Berry. M. J. A & Linoff. G., Data mining techniques: For marketing sales and customer support, New York, NY :John Wiley & Sons, 2004.
- [3] 김정자, 이도현, 데이터마이닝 기술 및 연구동향, 정보과학회지. 제16권 9호, 1998.
- [4] 이선영, 이종연, 유비쿼터스 컴퓨팅 환경에서 데이터 마이닝을 이용한 서비스 조합, 정보처리학회논문지D, 제13-D권 제4호, 2006.
- [5] Keita Fujii and Tatsuya Suda, Dynamic Service Composition Using Semantic Information, ICSSOC04, pp.15-19, November, 2004.
- [6] 김근형, 오경훈, 데이터마이닝 시스템의 커스터마이징을 위한 탐색전 연구, 한국지능정보시스템학회 추계학술대회논문집, 2004.6.
- [7] 석보현, 허의남, 정용환, 한국인터넷정보학회 정기총회 및 추계학술발표대회 논문집, 제7권 제 2호, 2006.11.
- [8] 이재식, 조유정, 한국지능정보시스템학회 추계학술대회 논문집, 2006.11.
- [9] Clifton. C., Kantarcioglu. M., Vaidya. J.(2004): Defining Privacy for Data Mining. Data Mining: Next Generation Challenges and Future Directions, AAAI Press, p.255-272.
- [10] Pinkas, B.: Cryptographic Techniques for Privacy-preserving Data Mining., ACM SIGKDD Explorations 4 ,2003.