

데이터마이닝을 기반으로 한 웹 전자상거래 서비스 설계

천린* · 김철원**

Design of E-Commerce Service on The Web Based on Data Mining

Lin Chen* · Chul-Won Kim**

요약

전자상거래의 추진력은 점점 강해지고 있으며 전자 상거래 간의 경쟁은 더욱 치열 해지고 있다. 이러한 데이터를 효과적으로 분석하고 합리적으로 사용하는 방법은 다양한 전자상거래 회사의 경쟁에서 중요한 포인트가 되었다. 본 논문에서는 데이터 마이닝 기술을 사용하여 대규모 웹 데이터베이스에서 중복 데이터를 필터링하고 유용한 데이터를 추출한 다음 다양한 관점에서 분석하여 이 데이터를 전자 상거래 웹 사이트에 합리적이고 효과적으로 적용한다.

ABSTRACT

The momentum of e-commerce is growing stronger and now, the competition between e-commerce is becoming more and more fierce. In the competition of various e-commerce companies, how to effectively analyze and rationally use these data has become a key point. This paper will use data mining technology to filter out redundant data from large Web databases, extract data that is useful to us, and then analyze them from different perspectives to apply this data reasonably and effectively to our e-commerce website.

키워드

Web Data Mining, E-Commerce, Data Preprocessor
웹 데이터 마이닝, 전자 상거래, 데이터 전처리

1. 서론

인터넷 기술의 발달과 전자 상거래의 보급에 따라 전자 상거래는 점차 우리 삶에 통합되고 전자 상거래의 발전은 전통적인 비즈니스 모델의 부진한 개발에 비해 눈에 띄게 되었다. 전자 상거래는 인터넷을 주제 및 정보 기술로 사용하는 현대적인 비즈니스 모델이다[1]. 전통적인 쇼핑 방식과 비교할 때 시간과 돈이

들기 때문에 대중의 소비 환경이 바뀌어 사람들이 쇼핑의 목적을 달성하기 위해 손을 움직일 필요가 없으므로 대부분의 사람들이 선호한다.

전자 상거래 사용자의 증가에 따라 전자 상거래 웹 사이트의 수가 증가하고 있으며, 웹 사이트 수가 증가함에 따라 고객에게 더 많은 선택권이 주어지며 고객에게 더 많은 어려움이 생겨났다. 금액이 더 크고 각 사용자

* 호남대학교 컴퓨터공학과(627504441@qq.com)

** 교신저자 : 호남대학교 컴퓨터공학과

• 접수일 : 2020. 05. 27

• 수정완료일 : 2020. 07. 06

• 게재확정일 : 2020. 08. 15

• Received : May. 27, 2020, Revised : Jul. 06, 2020, Accepted : Aug. 15, 2020

• Corresponding Author : Chul-Won Kim

Dept. of Computer Engineering, Honam University,

Email : cwkim@honam.ac.kr

에게 필요한 정보가 다르다. 그런 다음 유형과 구조가 다른 거대한 양의 정보에 직면 할 때 필요한 정보를 찾는 방법이 주된 문제가 되며 데이터 마이닝이 문제를 해결하는 핵심 기술이다. 데이터 마이닝은 대용량의 데이터 속에서 유용한 정보를 발견하는 과정이며, 기대했던 정보뿐만 아니라 기대하지 못했던 정보를 찾을 수 있는 기술을 의미한다. 데이터 마이닝을 통해 정보의 연관성을 파악함으로써 가치 있는 정보를 만들어 의사 결정에 적용함으로써 이익을 극대화시킬 수 있다[2].

II. 웹 데이터 마이닝

2.1 웹 데이터 마이닝 정의

가장 기본적으로 데이터 마이닝은 많은 양의 데이터로부터 지식을 발견하는 과정이라고 정의할 수 있다. 데이터마이닝은 새로운 원리이기보다는 몇 가지 서로 다른 과학적 원리들이 교차하는 새로운 접근방법이다. 데이터마이닝은 통계, 인공지능, 기계학습, 경영과학, 정보 시스템 등을 포함하는 여러 원리들의 능력들을 체계적이고 협력적인 방법으로 활용한다[3]. 경제 분야에서 데이터 마이닝을 적용하는 것은 현재 관심과 학자들의 뜨거운 주제가 되었다. 데이터 마이닝 기술을 사용하여 네트워크에서 생성 된 대용량 데이터를 분석하고 처리하면 기업이 고객을 분류하고 제품 판매를 예측하며 개발 계획을 세울 수 있다.

2.2 웹 데이터 마이닝 종류

다양한 마이닝 개체에 따라 WEB 기반 데이터 마이닝을 세 가지 범주로 나눌 수 있다: 웹 콘텐츠 마이닝(WebContent Mining), 웹 사용 마이닝(WebUsage Mining)과 웹 구조 마이닝(WebStructure Mining).

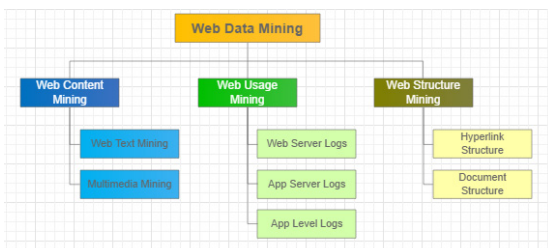


그림 1. 웹 데이터 마이닝 구성도
Fig. 1 Web data mining diagram

1) 웹 콘텐츠 마이닝(WebContent Mining)

웹 콘텐츠 마이닝은 웹 페이지에 저장된 콘텐츠로부터 웹 사용자가 원하는 정보를 빠르게 찾는 기법으로 검색엔진에 많이 사용된다.

웹 정보 자원에는 많은 종류가 있으며, 현재 WWW 정보 자원은 네트워크 정보 자원의 주요 구성 요소가 되었다, 그러나 사용자가 웹페이지를 통해 직접 액세스하여 사용할 수 있는 데이터 외에도 숨겨진 부분의 정보가 검색 될 수 없다, 이런 정보도 발견하고 조직해야 한다. 정보 자원의 성능 측면에서 볼 때 웹 콘텐츠는 텍스트, 이미지, 오디오, 비디오, 메타 데이터 및 기타 형식의 데이터로 구성되므로 웹 콘텐츠 마이닝은 멀티미디어 데이터에 대한 광업 유형이기도 한다[4].

2) 웹 사용 마이닝(WebUsage Mining)

웹 사용 마이닝은 또는 웹 로그 마이닝(WebLog Mining)이라고도 한다. 웹 콘텐츠 마이닝 및 웹 구조 마이닝과 달리 인터넷의 원본 데이터는 마이닝 개체로 사용되며 웹은 사용자와 네트워크 간의 상호 작용 중에 추출 된 두 번째 데이터를 사용한다. 이러한 데이터에는 웹 서버 액세스 레코드, 프록시 서버 로그 레코드, 사용자 등록 정보 및 사용자가 웹 사이트를 방문 할 때의 동작이 포함된다. 웹 사용 마이닝은 웹 사용자가 웹사이트에서 사용한 데이터를 통해 나타난 행위에 따라 그들의 특성과 성향을 추출한다[5].

3) 웹 구조 마이닝(WebStructure Mining)

웹 구조 마이닝 작업은 웹 페이지에서 이들 간의 조직 구조를 발견하는 것이다. 웹 구조 마이닝은 주로 웹 페이지 간의 링크 구조, 내부 페이지의 구조 및 URL 경로의 구조에 중점을 두고 있으며 웹 페이지의 구조는 복잡하며 사용자 요구에 따라 고유 한 디자인을 사용한다[6].

2.3 웹 데이터 마이닝 문제

1) 웹의 정보는 지속적으로 증가하고 변화하고 있으며, 각 데이터 소스의 구조와 정보는 다르므로 데이터 마이닝에 사용하려면 사이트 간 이기종 데이터의 통합을 연구해야하며 이러한 사이트의 데이터

만 연구한다. 모두 통합되어있어 사용자가 필요한 것을 쉽게 볼 수 있다.

2) 전자 상거래 웹 사이트의 데이터는 규모가 크고 복잡 할뿐만 아니라 특정 데이터 모델이 없으며 각 사이트의 데이터는 독립적으로 설계되며 데이터 자체에는 자체보고 및 동적 변동성이 있으므로 일반 데이터베이스의 데이터와 관련이 있다. 반대로, 그것은 완전히 구조화되지 않은 데이터 유형이다. 이 데이터를 처리하면 채굴의 어려움이 크게 증가한다.[7]

III. 웹 데이터 마이닝 모델 구축

웹 데이터 마이닝 문제에 대응하여 그림 2와 같이 데이터 마이닝 프로세스를 다음 모델로 설정했다. 먼저 웹 서버에서 소스 데이터베이스로 데이터를 변환한 다음 데이터를 데이터 웨어하우스로 사전 처리 한 다음 데이터 웨어 하우스에서 사전 처리한다. 데이터는 정규화되고 다른 처리 된 알고리즘은 처리, 변환 및 통합 된 데이터를 채굴하고 잠재적으로 유용한 패턴을 발견하는 데 사용된다. 이러한 패턴을 채굴 한 후에는 신뢰성과 효율성으로 패턴을 분석하고 최종적으로 분석 결과를 스테이션 관리에 적용하여 웹 사이트의 실제 경쟁력을 향상시킨다.

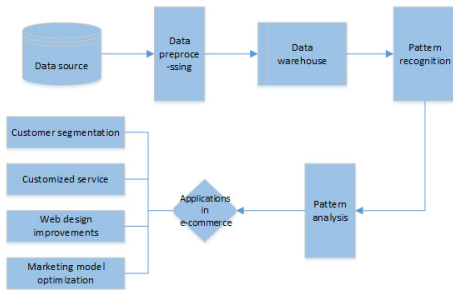


그림 2. 웹 데이터 마이닝 모델
Fig. 2 Web data mining model

3.1 데이터 소스 수집

데이터 소스는 주로 웹 페이지 데이터, 하이퍼링크 데이터 및 웹 사이트 사용자의 브라우징 레코드 및 트랜잭션 레코드이며, 데이터 수집은 주로 관련 데이

터를 수집하여 데이터웨어하우스에 저장 한 다음 데이터 분석 기술을 사용하여 데이터를 처리한다.

3.2 데이터 전처리

소스 데이터베이스의 원시 데이터의 양이 매우 커서 모든 데이터가 유용한 것은 아니므로 데이터 마이닝을 수행하기 전에 이 데이터를 사전 처리해야한다. 먼저 웹 로그에서 마이닝 목적과 관련이 없는 데이터를 삭제한 다음 웹 로그 레코드에서 필요한 데이터를 추출한 다음 추출 된 데이터의 기본 값을 피하기 위해 다른 기술적 방법을 사용하여 추출 된 데이터를 정제하다. 값, 특이치 및 노이즈도 후속 분석 프로세스를 방해하고 영향을 미치다. 그런 다음 정제 된 데이터는 다른 데이터웨어 하우스에 특정 형식으로 저장되어 이후 데이터 사용 효율성을 향상시킨다. [8] 그림 3은 데이터 전처리과정이다.

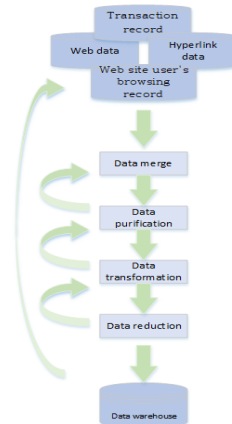


그림 3. 데이터 전처리과정
Fig. 3 Data preprocessing process

3.3 패턴 인식

이 단계에서 클러스터 분석, 분류 분석, 연관 규칙, 통계 분석 및 경로 분석과 같은 기술이 데이터웨어하우스의 사전 처리 된 데이터에서 추출되어 잠재적이고 효과적이며 사용 가능한 패턴을 추출한다.

1 군집 분석 : 군집 분석은 원본 데이터베이스의 데이터를 분석 및 비교 한 다음 유사성에 따라 여러 범주로 분류하는 것이다. 일반적으로 사용자 클러스터링과 웹 페이지 클러스터링으로 구분된다.

2 분류 분석 : 분류 분석은 소스 데이터베이스의 개별 데이터를 유형에 따라 다른 유형으로 지정하여 동일한 유형의 고객에게 유사한 서비스를 제공 할 수 있도록 한다.

3 연관 분석 : 연관 분석은 소스 데이터베이스에서 서로 다른 데이터 유형간에 숨겨진 연관 규칙을 찾는 것이다.

4 통계 분석 : 통계 분석은 가장 자주 방문한 웹 페이지, 각 웹 페이지 방문 시간 및 웹 페이지 경로를 수집하여 고객의 기본 정보를 얻는 것이다.

5 경로 분석 : 경로 분석은 빈번한 액세스 경로를 찾는 방법으로, 웹 서버의 로그 파일에서 클라이언트가 액세스 한 사이트 방문 횟수를 분석하여 빈번한 액세스 경로를 검색한다.

3.4 패턴 분석

패턴이 식별 된 후에는 이러한 모델의 신뢰성과 유효성을 분석하여 웹 사이트에서 이 지식을 웹 사이트에 합리적으로 적용하여 웹 사이트의 경쟁력을 향상 시켜야한다.[9]

IV. 전자상거래에서 응용

전자 상거래에서 웹 데이터 마이닝 기술을 적용하면 전자 상거래 기업의 경쟁력이 크게 향상되고 거대한 웹 페이지와 관련 링크 및 서비스 로그에 의미 있는 사용자 액세스 패턴과 잠재적 인 사용자 그룹 지식이 숨겨집니다. 데이터 마이닝을 최대한 활용하여 비즈니스 의사 결정 지원에서 핵심 데이터를 추출하고 웹 사이트 디자인을 개선하며 비즈니스 효율성과 관리를 향상시킬 수 있다.

4.1 고객 세분화

고객을 세분화함으로써 고객의 요구 사항을 보다 정확히 이해하고 파악할 수 있으므로 고객을 차별화하여 효과적으로 고객 충성도를 높이고 잠재 고객을 발굴 할 수 있다. 웹상의 고객 액세스 정보를 수집 할

때 분류 기술은 사용자에게 레이블을 붙이고 분류하는 데 사용되며 일반적으로 구매하지 않은 고객, 비정기 고객 및 자주 구매하는 고객의 세 가지 범주로 나눌 수 있다. 그림 4는 주별 사용자의 구매 빈도 분포도이다. 이 구매 빈도에 따라, 자주 구매하는 대상 고객이 정확하게 밀리고, 때때로 고객을 위해 온라인 프로모션이 제공되며, 아직 구매하지 않은 고객에 대한 사용자 리콜이 이루어진다.

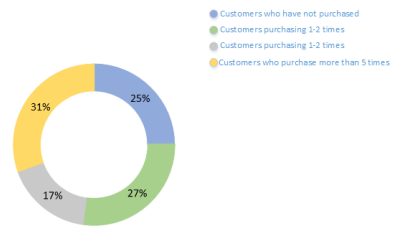


그림 4. 주별 사용자의 구매 빈도 분포도
Fig. 4 Weekly frequency distribution of users

4.2 맞춤형 서비스

생활수준의 향상과 변화하는 추세에 따라 고객의 요구 사항은 정적이 아니며 고객의 관심에 따라 유연하게 고객을 유치 할 수없는 경우 점차적으로 고객을 잃게 되므로 고객의 성격 서비스가 트렌드가 되어다. 그림 5와 같이, 고객의 액세스 로그 기록 정보 마이닝을 통해 고객의 브라우징 동작을 이해하고 고객의 관심 선호도 및 요구 사항을 파악한 다음 데이터 마이닝의 시퀀스 패턴과 결합하여 웹 페이지를 동적으로 조정하고 사전에 고객에게 제품 권장 사항을 제공함으로써 고객은 관심 있는 제품을 쉽고 빠르게 찾을 수 있다.

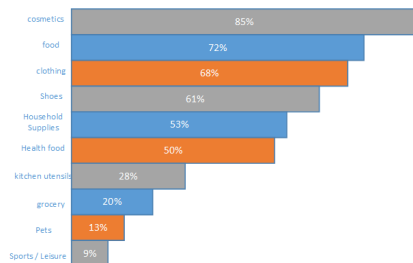


그림 5. 고객의 브라우징 행동 분포
Fig. 5 Distribution of browsing behavior of a customer

4.3 웹 디자인 개선

웹 페이지는 고객과 기업 간의 접촉에 대한 첫인상으로, 웹 페이지 디자인이 좋지 않으면 브라우징 시간과 웹 페이지에서 고객의 욕구가 직접적으로 줄어들므로 좋은 웹 디자인은 성공적인 전자 상거래 웹 사이트를 위한 것이다. 그렇다면 웹 페이지의 구조를 어떻게 최적화 할 수 있을까? 먼저 웹 로그 파일을 마이닝하여 사용자 액세스 페이지 간의 관련성을 확인하고 관련성 높은 페이지 간에 링크를 추가함으로써 사용자의 편의성을 향상시킬 수 있다. 둘째, 경로 분석 기술을 사용하여 웹 페이지에서 액세스 빈도가 높은 경로를 판단한 다음 고객 관심 분야에 대한 제품 정보를 이 페이지에 표시함으로써 사용자 브라우징의 복잡성을 줄이고 고객 체류 기간을 늘릴 수 있다. 결국 매출 증대의 목표가 달성된다.

4.4 마케팅 모델 최적화

모든 고객은 자신의 구매 습관을 가지고 있으며, 다른 고객이 동일한 것을 사더라도 구매 동기는 다를 수 있다. 그림 6과 같이, 일부는 가격에 주의를 기울이고, 일부는 품질에 주의를 기울이고, 일부는 외관에 주의를 기울이며, 따라서 일반적으로 시간과 공간에서 고객 구매 행동의 웹 마이닝을 수행 할 수 있다. 시간의 측면에서 보면, 고객의 액세스 규칙은 상품의 액세스 및 판매를 분석하여 얻을 수 있고 공간에서 특정 제품을 구매 한 고객에 대한 도메인 이름 분석을 수행하고 제품 판매가 지역적인지 여부를 확인할 수 있다. 다양한 유형의 마이닝 방법을 통해, 고객의 온라인 구매 습관을 이해하고, 데이터에 숨겨진 패턴, 연관성, 규칙 및 경향을 찾고 그 중에 규칙을 확인하고 교차 네트워크 마케팅, "일 대 일" 마케팅, 빈도 마케팅, 회원 마케팅 및 기타 마케팅 모델의 효율성을 향상시킨다[10].



그림 6. 고객의 쇼핑 동기 분석
Fig. 6 Analysis of customers' shopping motivation

V. 결 론

전자 상거래는 미래의 비즈니스 개발에서 피할 수 없는 추세이며, 기업의 비즈니스 활동 방식을 변화 시킬 뿐 아니라 사람들의 소비 방식을 변화시키다. 웹 마이닝은 데이터 마이닝의 최첨단 기술로, 웹 데이터 마이닝 기술을 사용하여 전자 상거래 웹 사이트의 다양한 데이터를 마이닝하고, 고객의 구매 습관을 분석하며, 관련 지식 패턴을 발견하여 전자 상거래 웹 사이트 리소스를 더 많이 구성 할 수 있다. 합리화는 많은 양의 데이터에 숨어있는 귀중한 정보를 찾고, 웹 사이트 페이지 디자인을 개선하고, 고객을 분류하고 잠재 고객을 찾고, 기존 고객에게 개인화 된 서비스를 제공하여 기업의 경쟁력을 효과적으로 향상시킬 수 있다.

References

- [1] F. Du, "Analysis of E-commerce Website Architecture Based on Web Data Mining" *E-business January*, Beijing, China, Feb. 2010, pp.69-70
- [2] Computational terminology compilation committee, *Dictionary of computer. internet. information thchnology terms*. Seoul: Iljinsa, 2007.
- [3] Dursun Delen, "Real-World Data Mining", England, 2016
- [4] C.Tu, M.Lu and Y.Lu, "Research on Web Content Mining Technology [J]." *Computer Application Research*, Beijing, China, Nov. 2003, pp. 5-9
- [5] CNI, "Unstructured Data Mining - Web Mining." Naver Blog, 2018
- [6] Z. Zhou and J. Zhu, "Research on Data Excavation Algorithm and Comprehensive Review," *Computer Engineering and Design*, Beijing, China, Sept. 2005, pp. 2304-2307.
- [7] M. Lee and K. Yoo, "Implementation of a Platform for the Big Scientific Data Transfers," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 13, no. 4, Aug. 2018, pp. 881-886.
- [8] E. Jung, C. Kim, and J. Park, "Design of Climate

Change Vulnerability Assissment Database System for Heat Wave and Droughr,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 13, no. 4, Aug. 2018, pp. 813-818.

- [9] M. Kim, “Scientometric Analysis Through Centrality Analysis of Graph for LINKAGE Relation of Keyword for Elder’s Rehabilitation and Healthcare,” *J. of the Korea Institute of Electronic Communication Sciences*, -vol. 14, no. 2, April. 2019, pp. 447-452.
- [10] C. Tan, “Analysis of the Relationship Between E-commerce Management and Web Data Mining Technology,” *Intelligence Journal*, Xian, China, Nov. 2006, pp. 9-12.

저자 소개



천린(Chen Lin)

2013년 대원민족대학교 컴퓨터공학과 졸업(공학사)
2016년 호남대학교 대학원 경영학과 졸업(경영학석사)

2020년 호남대학교 대학원 컴퓨터학과 졸업(공학박사)

※ 관심분야 : 데이터 마이닝, 인공지능



김철원(Chul-Won Kim)

1997년 광운대학교 컴퓨터공학과 졸업(공학박사)

1988년~현재 호남대학교 컴퓨터공학과 교수

※ 관심분야 : 멀티미디어 정보처리 및 응용
멀티미디어 정보검색 XML응용