

시간에 따라 변하는 사용자 관심도 학습

권현철⁰, 박영택
승실대학교 컴퓨터학과

bytecode@multi.soongsil.ac.kr, park@computing.soongsil.ac.kr

Profile Learning for Concept Change Over Time

Hyun-Chul, Kwon⁰, Young-Tack, Park
Dept. of Computing, Soongsil University

요 약

근래에 들어서 인터넷의 발전에 따라 사용자의 정보 검색 및 정보 서비스 이용에 대한 수요량이 많아지고 있으며, 이와 동시에 사용자 개인마다 적합하지 않은 정보에 대한 검색 시간과 서비스 이용에 대한 비용이 늘어나고 있다. 이에 따라서 사용자가 인터넷을 이용하면서 일어나는 행위들에 대한 정보를 수집하고, 이를 학습하여 생성한 사용자 프로파일을 기반으로 사용자 개인마다 맞춤형 적합한 정보를 제공하는 개인화 서비스가 늘어나고 있다. 본 논문에서는 사용자의 여러 행위에 대해 비 감독 학습 방법인 클러스터링을 이용하여 사용자 관심 클러스터를 생성, 사용하여 기존의 사용자 프로파일 학습에서 간과하고 있는 시간에 따라 변화하는 사용자의 관심에 대한 변화를 탐지하고, 변화하는 사용자의 관심 이동 형태에 따라 이를 사용자 프로파일을 생성하는 학습에 적용할 수 있도록 하는 방법을 제시하므로 해서 기존의 개인화를 위한 사용자 프로파일 학습 방법보다 진보한 학습 방법을 지닌 시스템 모델을 제시하려 한다.

1. 서 론

근래에 들어, 인터넷을 통해서 사용자에게 제공되어지는 정보의 수가 기하급수적으로 증가함에 따라서 사용자 개개인 원하는 정보를 정확하고 신속하게 검색하는데 많은 어려움이 있다. 이에 따라서, 수많은 정보 중에서 사용자의 관심에 따라서 정보를 제공받을 수 있도록 하는 정보 여과 시스템에 대한 요구가 있어왔다.

최근에 이런 요구는 개인화(Personalization)라는 방법을 통해서 사용자에게 맞춤형 정보를 제공하는 형태로 연구가 진행되어지고 있으며, 인터넷을 이용하는 사용자의 행위를 통해서 얻어진 정보를 수집하여 사용자 개개인의 프로파일(Profile)을 작성하고 이를 기반으로 학습한 결과를 가지고 정보를 여과(Filtering)하여 제공하고 있다.[1] 대부분의 개인화된 정보를 제공하는 시스템들은 정보를 제공하기 위해서 먼저 사용자의 관심에서 비롯된 행위에 대해 모니터링하여 정보를 추출하고, 이런 정보 중에서 사용자의 관심을 표현하는 프로파일로 사용될 주요 특성 선택(Feature Selection)한다. 이렇게 선택된 특성을 이용하여 구성된 프로파일을 통해서 개개의 사용자에게 맞춤형 정보를 제공하기 위한 학습을 재귀적으로 수행하고, 수행 결과를 이용해서 사용자에게 적합한 정보를 제공하고, 이에 대한 사용자의 관심에 대한 판단을 사용자의 피드백(Feedback)을 통해서 얻어진 정보로 재학습에 이용하는 방법을 사용하고 있다.

본 논문은 기존의 개인화를 위한 프로파일 학습 방법에서 간과하고 있는 시간에 따른 사용자 관심 변화(Concept Change)에 대해서 연구한 내용을 소개하고, 실제적으로 변화하는 관심, 즉 관심 이동(Concept Drift)에 대한 탐지 및 학습에의 적용에 연구의 목적을 둔다. 또한, 실제로 비감독 학습(Unsupervised Learning)방법인 클러스터링(Clustering)을 이용해서 사용자 행위에 대한 관심을 클러스터로 만들어 관심을 표현하고, 이를 이용해서 관심 이동에 대한 탐지를 수행하고, 탐지된 결과를 이용해서 프로파일 학습에 적용하여 시간에 따라 변화하는 사용자 관심도 학습을 할 수 있는 시스템 모델에 대해서 개인화

에서의 기대 효과를 서술한다.

본 논문에서 1장은 논문의 배경 및 동기에 대해서 서술하였다. 2장에서는 본 논문의 관련연구에 대해서 알아보고, 3장에서는 본 논문에서 제시하고 있는 학습 방법을 지닌 시스템 구성에 대해서 소개한다. 4장에서는 본 논문에서 목적으로 두고 있는 개인화 서비스를 위해서 시간에 따라 변하는 사용자 관심도 학습 시스템 모델의 구조를 보인다. 마지막으로 5장에서는 결론 및 향후 과제에 대해서 서술하여 맺는다.

2. 관련연구

2.1 개념 기반의 클러스터링

클러스터링이란 무질서한 데이터들 중에서 서로 연관이 있다고 판단되는 데이터들을 서로 묶어 주는 방법을 말한다. 클러스터링 기법은 무질서한 데이터들의 상호 연관관계를 정의하고 이를 통하여 보다 체계적이고, 정형화시켜 데이터들을 군집화시킬 수 있다. 이러한 클러스터링 기법은 인터넷에서 문서를 브라우징하는 특정 사용자의 프로파일을 추출하는 기본 방법으로 자주 사용된다. 사용자가 브라우징한 문서들을 클러스터링 기법으로 분류하고, 각각의 문서 분류 내에서 사용자의 프로파일을 추출해 내는 것이다. 클러스터링을 수행하는 기법에는 개념적 클러스터링(Conceptual Clustering)방식이 있는데, 이는 실세계 인간의 분류 방식을 모델로 삼아 그와 유사한 방식으로 입력 데이터들을 분류하는 것을 말한다. 클러스터링 방식에서는 입력데이터들의 속성과 정보들을 모두 받아 한번에 처리하는 일괄처리(Lazy)방식과 입력 데이터들을 하나씩 입력받아 유사한 데이터들을 클러스터링하는 점진적(Incremental) 방식이 있다.[2]

2.2 관심 이동

시간에 따라 변화하는 사용자의 관심을 파악하는 관심 이동(Concept Drift)에 대한 연구는 사용자 프로파일 구축 기술로서 연구되어져 왔다. Mitchell 등은 회의 시간의 조정에서 사용자 개개인의 스케줄 성향을 학습하여 시간을 배정하는 회의시

간 스케줄링을 위한 소프트웨어 도우미를 기술한바 있다.[3] Billisus와 Pazzani는 사용자의 흥미의 변화를 적용할 수 있도록 관찰 시간에 따라서 적용 속도의 차이를 두어 두 가지의 사용자 모델인 단기적 모델과 장기적 모델을 사용한 NewsDude라는 지능 에이전트를 기술하였다.[4] Widmer와 Kubat은 숨겨진 정황(Hidden Context)과 관심의 이동을 가진 환경에서는 효과적인 학습을 위해서 명확히 나타나지 않는 정황 변화를 탐지해 낼 수 있고, 변화를 빠르게 복원하고, 새로운 정황에 대해서 기존 관심의 조절을 가능케 하며, 다시 나타나게 될 관심을 위해서 과거 경험을 사용할 수 있는 알고리즘을 필요로 한다고 기술하였다.[5] Ingrid와 Klinenberg는 시간에 따라 변화하는 사용자의 관심에 대해서 그에 따라 데이터의 윈도우 사이즈(Window Size)를 변화시키는 사용자의 과거 정보에 대한 기간을 유동적으로 결정하는 자료 관리 접근법(Data Management Approach)의 연구를 통해서 적응형 윈도우 크기 관리(Adaptive Window Size Management)를 이용한 관심 이동을 적용시킬 수 있는 방법을 연구해왔다. 관심 이동에 대한 지시기(indicator)로서 사용된 성능 측정(Performance Measures) 방법은 관심 변화 탐지를 위해 분류기 성능을 3가지 측정값인 정확성(Accuracy), 정밀성(Precision), 재현성(Recall)으로 계산하여 관심 이동의 형태를 안정적인 관심(Stable Concept), 뜻밖이고 급진적으로 변화하는 관심(Concept Shift), 점차적으로 서서히 변화하는 관심(Concept Drift)으로 구분 지어서 윈도우 사이즈를 결정하는 알고리즘이 수행한다.[6] 관심의 형태에 따라 데이터의 윈도우 사이즈를 조절하여 학습에 적용하도록 하므로 해서 사용자의 관심에 접근한 분류를 할 수 있도록 한다.

2.3 개인화

개인화(Personalization)에 대한 연구는 인터넷 환경에서 사용자에 대한 정확한 이해를 통해, 직접적이고, 개인화된 관계를 형성하여 사용자의 요구를 최대한 만족시키기 위해서 진행되어왔다. 이를 위해, 카네기 멜론 대학(CMU)에서 구축한 웹 에이전트인 Personal WebWatcher의 연구를 수행하였다. Personal WebWatcher는 웹 브라우저 상에서 실행되면서 사용자의 행동을 모니터링하여, 개인 사용자에 대한 적응력(Adaptiveness)을 가지는 웹 에이전트 시스템이다. Personal WebWatcher는 사용자의 특성을 학습하는 방법으로 비감독(Unsupervised) 학습 방식으로 사용자 관심 문서의 전처리 작업과 학습 작업에는 추출된 키워드를 벡터 테이블로 생성, 이를 기반으로 TFIDF 및 베이저안 확률(Bayesian Probability)을 적용하여 사용자 프로파일이 구축된다.[7] 이로써 사용자 관심 문서를 시스템이 자동으로 추측하기 때문에 사용자에게 편리함을 준다는 이점이 있지만, 학습에 입력된 관심 정보의 신뢰성이 낮다는 단점이 있다.

3. 시스템 구성

본 논문에서 목적으로 두고 있는 시간에 따라 변화하는 관심 이동에 대한 탐지 및 학습에의 적용에 초점을 맞추어 볼 때, 기존의 시스템들 즉, 개인화 서비스를 위한 프로파일 학습을 수행하는 시스템들로 구성되어있다. 일반적으로 모니터링된 사용자의 행위에 대한 정보를 수집, 이를 분석하여 빈번히 발생하는 주요 키워드에 가중치를 부여하는 방식을 사용한다. 이런 키워드들은 벡터 테이블 형식으로 작성하여 예를 들어, TFIDF 방식 등을 적용하여 사용자의 관심 정보의 특징으로 추출, 이를 개인화된 정보 제공 시에 사용되어지는 사용자의 프로파일을 구축한다. 그리고, 제공된 개인화 정보에 대해서 개개의 사용자마다 결과에 대한 피드백을 받아 추후 학습에 이용하여 학습 결과 향상을 줄 정보로 사용되어진다.

본 논문에서 제안하는 모델은 대부분의 개인화 정보를 제공하는 시스템에서 간과하고 있는 시간에 따라서 변화하는 사용자 관심을 학습에 적용할 수 있도록 하였고, 이에 따라 다음을 고려하여 구성한다.

첫째, "사용자의 행위를 일으키는 것이 무엇인가?"이다. 인터넷을 통해서 여러 가지 작업을 하는 사용자의 행위는 사용자가 어떠한 목적이나 관심을 두고, 이를 이루기 위한 행동입에서 착안해야한다. 즉, 행위 발생은 사용자의 관심에 대한 표현인 것이다. 그러므로, 사용자의 관심에 근접된 표현을 할 수 있는 방법이 필요하다. 본 논문에서는 사용자의 행위에 대해서 모니터링 하여 수집된 문서들을 비감독적 학습을 수행하는 개념적 클러스터링 방법을 이용하여, 행위를 일으킨 원인인 관심을 클러스터 형태 즉, 관심 클러스터를 생성하여 표현하도록 한다. 생성된 관심 클러스터들은 사용자가 가진 관심에 대한 숨겨진 상황이나 관계들을 얻어 낼 수 있도록 관찰의 대상이 된다.

둘째, "사용자의 관심 변화는 어떻게 이루어지는가?"이다. 실생활에서 시간의 흐름에 따라 또는 사용자 주변 환경의 변화에 따라 과거 관심을 가지고 있던 사물이나 정보는 현재나 미래에는 관심 밖의 것들이 되어질 수 있다. 즉, 시간에 따라서 사용자의 관심은 이동하게 되는 것이다. 전에 가지고 있던 정보에 대한 관심도는 낮아지고, 반면 최근에 일어난 일들에 대한 관심도는 높아진다. 관심도가 낮아지는 과거 정보는 현재 관심의 변화에 의해서 필요 없는 정보가 되어 제거할 수 있다. 그러나, 제거된 정보는 사용자가 가지고 있는 일정한 관심일 수 있음을 배제할 수 없다. 실제 사용자는 이런 정보에 다시 관심을 두어 사용할 수도 있음을 말한다.

본 논문에서는 시간에 따라서 변화하는 관심을 탐지하기 위해서 클러스터링 방법을 사용하여 관심을 나타내는 클러스터를 생성할 때, 사용자가 흥미를 가졌던 정보의 내용(Content-based)과 시간에 따라 입력된 정보의 시간 속성(Temporal Property)을 고려하여 클러스터를 구성한다. 이때, 각 정보들은 현재 시점에서 정보들이 가진 시간 속성에 따라 클러스터를 이루는 정보들의 가중치를 계산하는 평가함수를 통해 클러스터를 구성하게 된다. 그러므로, 시간에 따라서 클러스터가 가지는 속성의 중심이 이동하게 되는데, 이를 측정하여, 그 변화량에 따른 관심 이동을 안정적인 관심(Stable Concept), 뜻밖이고 급진적으로 변화하는 관심(Concept Shift), 점차적으로 서서히 변화하는 관심(Concept Drift)으로 구분하여, 관심 이동의 형태에 따른 관심 클러스터를 생성하기 위한 시간 윈도우(Time Window)를 변화시킨다. 관심 변화에 따라서 윈도우의 사이즈는 안정적 관심일 때는 그 사이즈를 서서히 증가시켜 새롭게 들어온 정보를 포함해서 클러스터를 재구성하고, 뜻밖이고 급진적으로 변화하는 관심은 윈도우 사이즈를 작게 하여서 최근에 들어온 정보들을 중심으로 클러스터를 재구성한다. 그리고, 점차적으로 서서히 변화하는 관심은 윈도우 사이즈를 조금씩 줄여가도록 해서 관심의 이동이 안정적인 상황이 될 때까지 최근의 정보만으로 구성되도록 클러스터를 재구성한다. 또한, 클러스터가 가진 속성의 변화는 사용자의 관심 변화를 일으킨 최근 정보들과 시간이 지남에 따라 오래되어 관심에 멀어진 정보로 나누어 볼 수 있는데, 이때 대부분의 시스템들이 오래된 정보들은 잊어버리고, 최근 정보들을 이용해서 관심 클러스터를 만든다[8]. 그러나, 이렇게 오래된 정보를 단순히 잊는 경우에는 장기적인 사용자의 관심을 무시하게 되는 것이다. 따라서, 사용자의 장기적인 관심(Long Term Concept)을 고려하여, 이로 인해서 단기적인 관심(Short Term Concept)에 도움을 줄 수 있도록 하였다. 즉, 사용자는 갑작스럽게 예전에 보였던 관심을 일시적으로 나타낼 수 있는데, 이를 수용할 수 없는 단기적인 관심을 위해서는 장기적인 기억(Long Term Memory)방식을 통한 장기적 관심을 이용하여 수용

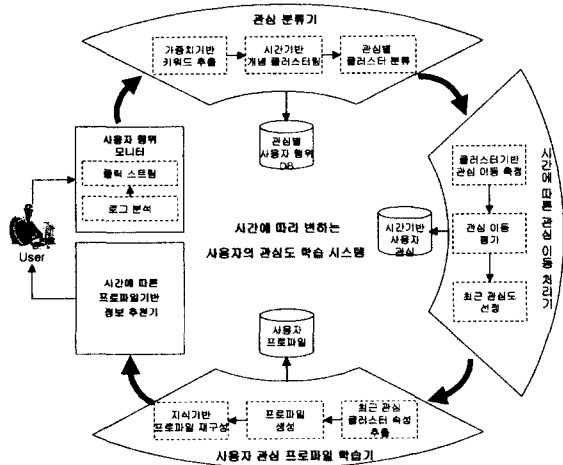
할 수 있도록 하였다. 따라서, 장기적인 기억을 통해서 기억되는 것은 많은 양의 정보들을 기억하는 것이 아니라, 과거에 가졌던 관심 클러스터의 대표 속성들을 저장하여 새로운 정보가 들어올 경우, 과거 관심과의 유사도를 통해서 일시적인 정보에 대해서 해결한다.

셋째, "시스템에서 생성해 내는 관심과 사용자가 가지고 있는 관심의 차이는 무엇인가?"이다. 일반적인 시스템에서 사용자 행위들을 분석하여 만들어지는 관심 분류기가 생성되면, 이를 통해서 관심에 대한 분류를 하게 된다. 그러나, 이런 분류가 인간의 실제 관심과 비교해볼 때 차이를 보이게 되는데, 이는 실제로 숨겨진 관계나 정황에 대해서 탐지해 내지 못하기 때문이다. 예를 들어, 온화한 날씨엔 시베리아와 중앙 아프리카에서의 의미가 다르다. 같은 정보에 대해서 유사한 프로파일 가지는 사용자마다 다른 관심을 보일 수 있다.

본 논문에서는 이런 점을 고려하기 위해서 클러스터링을 통해서 만들어진 관심 클러스터를 사용자 프로파일로의 적용에 있어서 같은 형태의 관심 클러스터이더라도 사용자들마다 그 의미를 달리 한다는 점을 착안하여 사용자에 대한 지식을 기반으로 프로파일 달리 생성을 하였다. 위와 같이 구성된 사용자 프로파일을 기반으로 한 시스템을 바탕으로 개인화 서비스 제공을 하여, 기존에 사용자 프로파일을 이용한 시스템보다 정확성을 높일 수 있다.

4. 시스템 구조

본 논문에서 제안하는 시간에 따라 변하는 사용자 관심도 학습 시스템 모델의 전체 구조는 [그림1]과 같으며, 주요 모듈에 대한 설명은 다음과 같다.



[그림1] 시간에 따라 변하는 사용자 관심도 학습 시스템 구조

4.1 사용자 행위 모니터

사용자가 현재 관심을 보인 문서에 대한 수집을 위해서 사용자 로그와 클릭 스트림 방식을 이용해서 모니터링하고, 이를 관심 클러스터를 생성하기 위해서 사용하기 위해서 데이터베이스에 저장하도록 수행한다.

4.2 관심 분류기

사용자 행위로 수집된 문서를 기반으로 개념적 클러스터링을 점진적으로 수행한다. 이때, 사용자가 관심을 가진 문서마다 시간적인 속성 값을 가지고 있으므로, 이를 기반으로 문서의

가중치를 달리 적용하여 사용자의 최근 관심도에 밀접하게 접근하도록 수행한다. 이 점은 본 논문에서 제시한 시스템 모델로 기존 클러스터링과 차별성을 가지는 한 부분이다.

4.3 시간에 따른 관심 이동 처리기

관심 클러스터가 생성되어지고 나면, 현재 관심 클러스터의 중심 값을 얻을 수 있다. 이렇게 얻어진 중심 값을 과거 값과 비교를 해서 변화량을 기록하고, 시간에 따른 변화 측정값을 이용하는 평가합수를 통해서 현재 관심 이동의 형태를 구분하고, 클러스터링을 수행하는데 사용되어진 정보들에 대해서 시간 윈도우를 조절하도록 하여, 추후에 클러스터링 진행에 영향을 주도록 한다. 따라서, 최근 사용자의 관심을 알고있는 정보를 얻을 수 있다. 이 같은 방법은 기존 시스템에서 지속적으로 입력 데이터를 이용하는 방식과 달리 시간에 따라 유동적으로 데이터를 관리하는 방식을 사용한다.

4.4 사용자 관심 프로파일 학습기 및 정보 추천

생성되어진 최근의 관심 클러스터 속성을 추출하고, 이에 따라 사용자 프로파일을 구축한다. 이렇게 구축된 사용자 프로파일은 사용자마다 보다 적합하도록 지식기반으로 프로파일 재구성한다. 이 프로파일 통해서 개인화된 정보를 추천하므로 기존의 시스템보다 사용자에게 정확성 높은 정보를 제공한다.

5. 결론

인터넷을 사용하는 사용자의 대부분은 시간이 흘러감에 따라서 관심을 가지고 있는 분야는 계속적으로 늘어나고 과거 가지고 있던 관심에서 변화를 보여 다른 것에 관심을 보인다. 실제 동적인 환경에 맞도록 하기 위해서는 기존 방식보다는 동적인 시스템을 요구가 이루어질 것이다. 따라서 본 논문에서 제시하는 시스템 모델은 시간에 따라서 사용자가 가진 관심 변화를 탐지하는 기능을 가지고 있어 유동적으로 대처할 수 있도록 되어있어, 개인화 서비스를 제공하는 사용자 관심에 대해 근접한 해결책을 제공할 것이다.

6. 참고 문헌

- [1] L.Dent, J. Boticario, J. McDermott, T. Mitchell, D. Zabowski, "A Personal Learning Apprentice", In Proceedings of the 11th International Conference on Machine Learning, July 1994
- [2] T. Mitchell, "Machine Learning", McGraw Hill, 1997
- [3] T. Mitchell, R. Caruana, D. Freitag, J. McDermott, D. Zabowski, "Experience with a Learning personal Assistant", Communications of the ACM 37.7 81-91, 1994
- [4] D. Billsus, M.J. Pazzani, "A Hybrid User Model for News Story Classification", In Kay J. (ed.), Proceedings of the Seventh International Conference on User Modeling, Springer-Verlag, 99-108, 1999
- [5] Gerhard Widmer, Miroslav Kubat, "Learning in the Presence of Concept Drift and Hidden Contexts", 1996
- [6] Ralf Klinkenberg, Ingrid Renz, "Adaptive Information Filtering: Learning in the Presence of Concept Drifts", 1998
- [7] Dunja Mladenic, "Personal WebWatcher: Implementation and Design", Technical Report IJS-DP-7472, October 1996
- [8] M. Maloof, R. Michalski, "Learning Evolving Concepts Using a Partial Memory Approach", Working Notes of the AAAI Fall Symposium on Active Learning, 1995