

사용자 맞춤 동적 지능형 환경을 제공하는 스마트 공간 프레임워크

(Smart space framework providing dynamic embedded intelligent information)

장서윤*, 강지훈**

(SeoYoon Jang, JiHoon Kang)

요약

스마트 공간은 주변 환경과 상호 작용하여 인간을 지원하는 기술이며, 내장된 동적 지능형 환경을 가지고 있다. 이 논문은 스마트 공간에서 사용자 맞춤 동적 지능형 환경을 제공하는 프레임워크를 제안한다. 사용자 맞춤 지능형 서비스를 제공하는 기존 연구에서는 사용자의 관심사를 명시적으로만 분석하고, 스마트 공간을 고려하지 않았다. 암시적 관심사 분석은 명시적 관심사 분석보다 사용자에게 흥미가 있을 수 있는 서비스의 제안이 가능하지만, 분석 시 명시적 관심사 분석보다 높은 성능을 요구한다. 그리고 스마트 공간은 공간의 정보들과 상호작용하여 유용한 정보를 얻을 수 있다. 연구에서 제안된 프레임워크는 스마트 공간에 적합하도록 근접 기반 사물 소셜 네트워크를 이용한다. 또한, 암시적 관심사 분석은 사물의 소셜 미디어 정보와 공간 정보를 이용하여 스마트 공간을 위한 지능형 정보를 제공한다. 그리고 스마트 공간 특성을 고려하여 정확도를 유지하면서 성능 저하를 막는 방법을 제안한다.

- 중심어 : 스마트 공간 프레임워크 ; 사용자 암시적 관심사 분석 ; 상황 맞춤 정보 ; 지능형 IoT 환경 ; 동적 지능형 환경

Abstract

Smart space is a technology that supports humans by interacting with the surrounding environment. Smart space has a built-in dynamic intelligent environment. This paper proposes a framework that provides user-customized dynamic intelligent environments in smart spaces. In the existing research that provides user-customized intelligent services, users' interests are only explicitly analyzed, and smart spaces are not considered. Implicit interest analysis can suggest a service that may be of interest to users rather than explicit interest analysis, but it requires higher performance than explicit interest analysis. Smart spaces can obtain useful information by interacting with information in the space. The framework proposed in the study uses a proximity-based social network of things to fit into a smart space. In addition, the implicit interest analysis provides intelligent information for smart spaces using the social media information and spatial information objects. In addition, we propose a method to prevent performance degradation while maintaining accuracy in consideration of the characteristics of the smart space.

- keywords : Smart Space Framework ; Context-specific information ; User implicit interest analysis ; Intelligent IoT environment ; built-in dynamic intelligent environment

I. 서 론

스마트 공간은 많은 양의 센서 정보와 상황 정보를 이용하여 동적 내장된 지능형 환경을 구축하는 것을 목적으로 한다. 예를 들어, 가정에서 에너지 사용을 최적화하거나 사용자에게 필요한 정보를 제공하여 사용자 공간의 서비스 환경을 풍부하게 한다 [1]. 특히, 스마트 공간의 사물은 공간의 상황 정보를 모니터링

및 수집하고 사용자의 환경과 상호 작용한다. 이 연구에서 제안된 스마트 공간 프레임워크는 모바일 소셜 네트워크 (MSN)를 이용하고, 공간을 이용하는 두 종류의 이기종 사물 에코 시스템을 고려한다. 첫 번째 사물은 스마트 공간에 진입하는 스마트폰이다. 스마트폰은 사용자의 정보를 가지고 있으며, 많은 센서 정보를 내장하고 있다. 다른 사용자인 고정 사물은 위치가 고정되어 공간의 정보를 가지고 있다. 우리는 공간에 있는 정보와 사물을 이용하여 상황에 맞는 서비스를 제공하므로 성능에 제약

* 정회원, 충남대학교 컴퓨터공학과 박사수료 장서윤

** 정회원, 충남대학교 컴퓨터융합학부 교수 강지훈

이 연구는 충남대학교 학술연구비에 의해 지원되었음.

이 있다고 가정한다. 우리는 스마트 공간에서 사용자에게 상황에 적합한 정보와 환경을 제공하는 사물을 위한 프레임워크를 제안한다. 특히, 동적 내장된 지능형 환경에서 사용자 관심사 분석 방법을 연구한다. 사용자 관심사는 일반적으로 소셜 미디어 데이터를 분석하여 작성된다. 소셜 미디어 데이터를 분석하면 사용자의 개인 선호도 및 시간에 따른 관심사 변동 등의 정보를 얻을 수 있으며 여러 연구에서 효과가 증명되었다[2-5]. 기존에 사물 환경에서 사용자 관심사를 적용하여 상황에 맞게 정보를 분산하여 분석 성능을 높인 연구인 LISO가 있다[6]. LISO는 사물에 맞는 정보만 소셜 네트워크에서 자료를 수집하여 각자 분석한다. 서비스가 필요한 경우 이미 분석된 데이터를 서로 교환하여 서비스 지역을 해결한다. 하지만 기존의 연구는 키워드 형식으로 지식 데이터를 저장하여 내용의 변경에 시간이 걸린다. 또한, 명시적 관심사 분석을 이용하여 서비스를 제공한다. 명시적 관심사는 수집된 자료에서 명확하게 보이는 정보만을 이용하여 만든 것이다. 일반적으로 사용자가 작성한 해시태그 그리고 직접 언급된 짧은 텍스트 등으로 얻을 수 있는 데이터이다. 그래서 사용자와의 관계적 정확성은 높지만, 암시적으로 사용자의 관심이 있을지도 모르는 관심사는 얻기 힘들다. 암시적 관심사는 명시적으로 언급하지 않았지만, 사용자가 관심이 있을 가능성이 있는 정보이다. 암시적 관심사를 적용하면, 기존에 발견하기 어려웠던 사용자의 관심사에 관련된 서비스를 제공할 수 있다[7]. 연구에서 제안된 암시적 관심사는 사물의 성능을 고려하여 분석을 진행한다.

본 연구에서는 기존의 문제점을 해결하고, 스마트 공간에 적합하게 설계된 암시적 관심사가 적용된 스마트 공간 프레임워크를 제안한다. 스마트 공간은 상황 정보에 대한 분석, 데이터의 교환 그리고 동적 지능화 분석이 포함되어야 한다. 우리는 먼저 스마트 공간 프레임워크를 설계한다. 프레임워크에는 데이터 교환, 소셜 네트워크 작성이 포함된다. 그 후, 동적 지능형 서비스를 제공하기 위한 스마트 공간을 고려한 암시적 관심사 분석 방법을 제안한다. 암시적 관심사 분석과 정보의 유연성을 높이기 위해 엔티티를 연결하는 그래프 방식으로 데이터를 저장한다. 지식 그래프는 기존 방법보다 의미적 연결성이 뛰어나다. 특히, 데이터의 수정 및 삭제에 좋은 성능을 보인다. 우리는 데이터를 wikidata의 엔티티 링커를 이용하여 지식 그래프로 구성한다. 암시적 관심사 분석은 혼합형 필터링과 Propflow를 이용한다. Propflow는 무작위 걷기 예측 모델 중 하나로, 무작위 전송을 기반으로 정보가 연결될 확률을 구하는 링크 예측 분석 방법이다.

LISO는 스마트폰이 고정 사물의 공간에 들어오면 서로 데이터를 교환한다. 예를 들어, 기존 연구에서 레스토랑 공간에 진입하였을 때 교환되는 데이터는 명시적인 데이터이다. 예를 들어, 사용자가 직접 언급한 돼지고기 비선호 등의 정보이다. 하지만, 수정된 스마트 공간 프레임워크를 적용하면 더욱 풍부한 서비-

스를 제공하며 사용자 환경을 수정한다. 예를 들어, 채식을 선호, 건강한 단맛을 선호 등의 정보가 제공된다. 그래서 고정 사물은 레스토랑의 최근 암시적인 관심사를 적용하여 건강한 단맛에 해당하는 설탕 대신 스테비아, 시럽류 등의 대체 감미료를 추가하는 메뉴를 제안한다. 또한, 센서 데이터의 암시적 관심사 분석으로 레스토랑의 좌석 배치 및 위치 안내 그리고 좌석의 조명 선호 조도 등의 환경을 조정한다.

본 연구의 구성은 다음과 같다. 2장에서는 관련 연구를 설명한다. 기존의 연구인 LISO를 소개하고, 소셜 미디어 데이터를 이용한 관심사 분석을 설명한다. 3장에서는 수정된 스마트 공간 프레임워크를 제안한다. 또한, 지능형 환경 정보 제공을 위한 암시적 관심사 분석 방법을 제안한다. 4장에서는 수정된 스마트 공간 프레임워크의 실험 및 평가를 진행한다. 마지막으로 5장에서 연구에 대한 전반적인 정리 및 결론을 제시한다.

II. 본 론

1. 관련 연구

가. LISO 프레임워크와 서비스

LISO[6]는 사물의 위치와 상황을 기반으로 소셜 미디어의 정보를 분석하여 지능형 서비스를 제공한다. 위치를 기반으로 소셜 미디어의 정보를 분석하면 분석 정보량을 줄일 수 있어 성능을 올리고 정확도를 향상할 수 있다. LISO에서 분석하는 정보는 상황 정보와 소셜 미디어 데이터이다. 특히, LIOS는 사물이 가진 상황 정보에 맞춰 동작하도록 설계되었다. 먼저 사용자가 들고 다니는 형태의 모바일 사물 LIOS는 위치가 지속해서 변경된다. 그리고 사용자가 소셜 미디어에 데이터를 생성하고 수정하는 데 사용한다. 이런 형태의 LIOS는 사용자의 정보가 자주 업데이트되는 상황 정보를 가진다. 소셜 미디어에서 광범위한 정보를 수집하는 것보다 사물에 적재된 정보를 분류하고 분석하는 것에 집중한다. 반면 위치가 고정되어 사용하는 고정 위치 사물 LIOS는 사물의 정보가 자주 변동되지 않는다. 상황 정보가 많이 변동되지 않아 소셜 미디어에서 정보를 자주 수집하여 저장한다. 즉, 상황 정보가 자주 변동되면, 사물이 생성하는 정보를 더 많이 사전 분석에 이용한다. 상황이 자주 변동되지 않으면, 사물이 필요로 하는 정보를 소셜 미디어 정보를 수집하여 사전 분석에 이용한다.

하지만, LISO는 명시적 관심사 분석만 제공하며, 사물의 상황 정보만 사용한다. 우리는 스마트 공간에 맞도록 센서 정보를 이용하여 사용자 관심사 분석을 하고, 사용자가 관심을 가질 수도 있는 암시적 관심사 분석을 적용하여 서비스를 더 풍부하게 만든다.

나. 사용자 관심사 분석

많은 사용자는 소셜 네트워크 게시물을 사용하여 실시간으로 발생하는 다양한 소셜 네트워크 이벤트 그리고 주제에 대한 감정과 견해를 표현한다. 이 내용은 소셜 네트워크를 사용자의 관심사에 대한 정보 소스가 될 수 있게 만든다[2,8,9]. 소셜 네트워크에서 텍스트 콘텐츠를 분석하여 관심사를 추출하는 연구는 꾸준히 진행되고 있다. 소셜 네트워크의 게시물은 제한 없이 자유롭게 게시할 수 있으므로 무제한에 가까운 용어를 가진다. 용어의 다양성을 분류하고 의미 표현을 풍부하게 적용하기 위해, 최근의 관심사 추출 작업에서는 KB(Knowledge Base)의 외부 지식을 활용한다[10,11]. KB를 사용한 기준의 연구는 일반적으로 명시적 관심사를 추출하여 서비스를 제공한다. 명시적 관심사는 사용자를 비교적 명확하게 모델링 할 수 있는 장점이 있다. 하지만, 명시적인 관심사 모델링은 지식 기반에 공식적으로 표시된 정보만 감지한다. 기준 연구의 대부분은 명시적 관심사 추출에 초점을 맞추고 있지만, 사용자의 암묵적인 관심사를 추출하는 데 집중한 연구도 있다. 기준의 한 연구는 사용자의 암시적 관심사를 추출하기 위해 그래프 기반 링크 예측기법을 제안하고 그 기능을 개념 증명으로 설명한다[12]. 제안된 기법은 주제의 관련 성과 사용자 간의 사회적 관계를 통합하여 표현 모델로 나타낸다. 또한, 주제 간의 연관성은 사용자의 내재된 관심을 드러낼 수 있는 또 다른 중요한 측면이다. 일부 연구자들은 주제간 연관성을 가진 사용자가 종종 일관된 관련 관심사를 가지고 있다고 주장한다[13]. 이를 바탕으로 사용자가 명시적인 관심을 보인 주제와 개념적으로 유사한 주제에 관심이 있을 가능성이 크다고 가정한다. 예를 들어, 미국 메이저 리그에 관심을 보인 사용자는 한국의 KBO리그에 관련된 정보에도 관심을 가질 수 있다. 실제로 2020년 미국 메이저 리그가 코로나19로 인해 연기되었다. 그러자 많은 미국의 팬들이 한국의 KBO리그에 관한 관심을 많이 드러내었다. 우리는 더 풍부한 서비스를 제공하기 위하여 스마트 공간 프레임워크에 암시적인 관심사를 적용한다.

2. 스마트 공간 프레임워크 설계

스마트 공간에서는 센서 정보를 이용하여 서로 데이터를 교환하고, 사용자에게 적절한 환경을 구성하는 서비스를 제공한다. 기준에 센서 정보를 제공하여 소셜 네트워크를 구성하고 분석하여 사용자에게 맞춤 서비스를 제공한 연구가 있다[14,15]. 우리는 그중에서도 근접 기반 소셜 네트워크를 고려한다. 지리적 근접성을 기본 필터로 사용하는 이 방법은 여러 방법으로 서로를 발견할 수 있다. 일반적으로 GPS 위치, 근거리 무선 통신 기술(WiFi Direct, Bluetooth 및 기타 근거리 무선 통신) 등을

사용한다. 우리는 모바일 소셜 네트워크(MSN)를 고려하여 프레임워크를 설계한다. 모바일 소셜 네트워크 네트워크(MSN)를 적용하는 혼합형 기능은 분산 스토리지를 고려하여 단거리 무선 기술로 데이터를 교환하기 때문에 인프라 전체 부하를 줄인다[15].

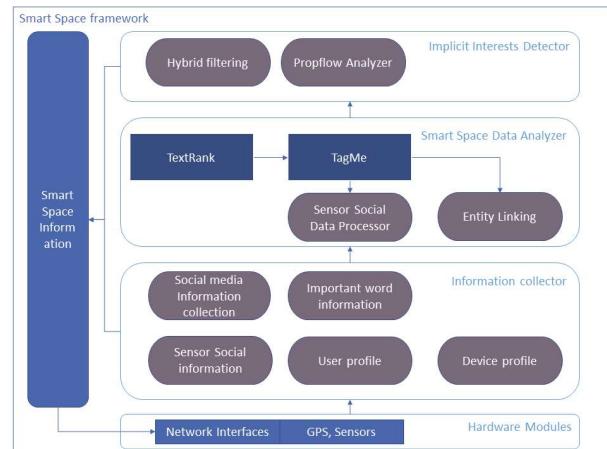


그림 1. 사용자 맞춤 지능형 정보를 제공하는 스마트 공간 프레임워크

연구에서 제안하는 사용자 관심사가 적용된 스마트 공간 프레임워크는 그림 1과 같다. 스마트 공간 프레임워크는 지식 데이터를 지식 그래프로 구성하기 위한 엔티티 링커와 암시적 관심사 추출을 하기 위한 암시적 관심사 분석기가 있다. 암시적 관심사 분석기 내부에는 시간의 경과에 따른 가중치 적용 분석기와 PropFlow을 위한 분석기가 있다. TextRank의 가중치 계산 및 이동할 확률값은 0.85를 사용하였다[16]. 수치를 변경하며 실험해 본 결과 PageRank에서 제안한 값에서 큰 차이를 보이지 않았다.

이 프레임워크는 스마트 공간에서의 센서 정보의 상황 분석을 고려하여 설계되었다. 상황 정보 분석에는 프로파일링, 데이터 수집 그리고 학습의 기능이 있다. 프로파일링에서는 사용자와 장치 정보를 수집한다. 사용자 프로필에는 사용자의 연락처, 방문한 장소 그리고 소셜 네트워크 데이터가 포함된다. 장치 프로필은 네트워크 인터페이스, 배터리와 센서를 포함한 장치의 기능 정보가 포함된다. 프로필은 모두 소셜 미디어 정보와 유사하게 지식 그래프 형식으로 가공되어 사용된다. 정보에는 시간과 위치에 정보가 반드시 포함된다. 그리고, 자원에 기여에 대한 시간적 사용을 사용자의 선호도로 기록한다. 특히, 프로필은 센서 소셜 네트워크 정보로 따로 가공하여 저장한다. 작성된 지식 그래프는 따로 저장하지만, 모두 소셜 미디어 정보를 수집과 암시적 관심사의 데이터로 사용된다. 사물 간 데이터 교환은 구조적 특징을 고려하여 XML 형식을 사용한다.

3. 스마트 공간의 동적 지능 정보 제공

기존의 LISO에서는 명시적 관심사를 이용하여 서비스를 제공한다. 우리는 사용자에게 더 적합한 정보를 제공하기 위해 LISO 공간을 확장한 스마트 공간에 암시적인 관심사를 적용하는 방법을 제안한다. 또한, LISO에서 사용한 상황, 위치 정보뿐 아니라, 프레임워크에서 작성된 센서의 소셜 네트워크 데이터도 암시적 관심사 분석에 이용하여 사용자에게 적합한 환경을 제공하기 위해 사용한다.

가. 지식 그래프형식으로 데이터 구조 변경

Wikidata의 엔티티 링커는 openTapioca를 사물에서의 지능형 서비스에 사용하기 위해 수정하였다. 사용자의 관심사를 찾기 위한 초기 인덱스는 기기의 검색기록, 소셜 네트워크 데이터 그리고 위치 기록을 이용하였다. 인덱스의 추출은 TextRank[16]를 이용하였다. 너무 넓은 범위의 엔티티 연결을 막기 위해 최대 5단계 이상의 연결을 넘어가지 않도록 관리한다. 이것은 기존 분류법의 범위를 기준으로 설정하였고, 적용을 위해 초기 인덱스 항목을 따로 표기하였다[17]. Wikidata로만 이루어진 초기 엔티티 연결 지식 그래프는 상황에 관련된 데이터와 세부 데이터가 부족하다. 부족한 세부 데이터는 명시적 분석을 이용하여 보완한다.

나. 암시적 관심사 분석

우리는 스마트 공간에서 사용자에게 상황에 적절한 지능형 서비스를 제공하기 위한 암시적 관심사 분석 방법을 제안한다. 암시적 관심사 분석은 기존의 명시적 관심사 분석보다 많은 성능과 시간을 요구한다. 먼저, 관심사를 표현하기 위한 기본 모델을 생성한다. 기본 모델은 지식 그래프 형식으로 변환된 데이터를 이용한다. G_z 는 일련의 주제 z 에서 활성 주제 간의 잠재적 관계를 나타내는 그래프이다. G_{uz} 는 사용자 관련 정보 u 에서 명시적으로 관찰 가능한 사용자-주제 관계 그래프이다. 기본 모델 그래프 G 는 각각의 그래프가 결합한 그래프이며, 공식 1과 같다.

$$G = G_z \cup G_{uz} \quad (1)$$

사용자-주제 관계 그래프 $G_{uz} = (V, E, g)$ 는 각종 무방향 그래프이다. $V = z \cup u$ 이며, 옛지 E 는 활성화 중인 주제 z 에 대한 사용자의 명시적 관심을 관찰하여 설정한다. 주제간 의미 연관성은 두 주제간 관련성을 의미한다. 두 개의 주제는 해당 주제를 구성하는 의미론적 개념이 유사한 정도를 계산하여 유사하다고 간주한다. 주제 유사성 분석을 위하여 콘텐츠 기반 필터링과 아이템 기반의 협업 필터링을 조합한 하이브리드 필터링 [18] 분석을 진행한다. 협업 필터링 방식은 계산 효율 저하와 룰 테일 문제로 인한 성능 저하를 문제가 발생한다. 그래서 사물에서의 단일 유사도 분석으로 사용하기에 적합하지 않다. 특히, 상황이 자주 변하는 모바일 기기에서는 종종 문제가 발생한다. 우리는 협업 필터링의 문제를 보완하기 위하여 콘텐츠 기반 필터링을 사용하는 하이브리드 방식을 적용하였다. 연구에서는 주제에 대한 사용자의 중복 기여도를 기반으로 관련성을 측정한다. 공식 2는 아이템 기반 협업 필터링에서 사용한 피어슨 유사도이다. 여기서 i, j 는 주제이다.

$$\text{pearson_sim}(i, j) = \frac{\sum_{u \in U_{ij}} (r_{ui} - \mu_i) \cdot (r_{uj} - \mu_j)}{\sqrt{\sum_{u \in U_i} (r_{ui} - \mu_i)^2} \cdot \sqrt{\sum_{u \in U_j} (r_{uj} - \mu_j)^2}} \quad (2)$$

콘텐츠 기반 필터링은 안정성을 위해 명시적 분석으로 얻어진 주요 키워드를 이용한다. 서비스를 제공하는 상황에서 다른 사물에게 받은 데이터는 먼저 콘텐츠 기반 필터링 분석으로 유사도 분석을 하고, 협업 필터링 분석으로 더욱 정확한 유사도를 계산한다. 이러한 방법을 이용하여 적절한 시점에 서비스를 제공할 수 있다.

기본 모델 그래프 G 를 구축하면, 사용자의 암시적 관심사를 분석하기 위해 사용한다. 사용자 u 가 명시적으로 관심을 표현하지 않은 각 주제 z 에 대한 암시적 관심의 분석은 PropFlow를 이용한다. 암시적 관심사를 분석하기 위한 여러 분석 방법이 있지만, 상황에 맞는 서비스를 제한된 환경에서 제공하기 위한 가장 적절한 암시적 관심사 분석은 메타 데이터를 이용한 예측 방법이다. 그리고 프레임워크에서 생성되는 지식 그래프 데이터는 항상 구조가 변동되는 이기종 그래프이다. PropFlow은 무작위 견기 예측 모델로, 서로 다른 형태의 그래프를 결합하는데 사용할 수 있다[19].

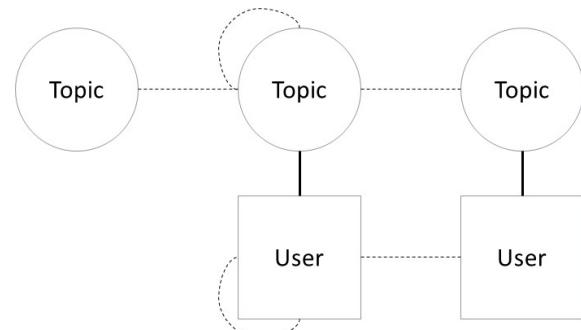


그림 2. 추출 가능한 메타 경로 구조

우리는 PropFlow을 사용하여 각 주제 z 에 대한 사용자 u 의 암시적 관심을 추출한다. PropFlow는 고정 길이 제약 내에

서 재귀적으로 정의된 모든 경로를 따라 무작위 전송을 하여 u 에서 z 로 정보가 흐를 확률을 구한다. 일반적으로 분석을 진행하면 비지도 학습 방법인 무작위 걷기로 생성된 메타 경로의 수는 최대 경로 길이가 기하급수적으로 증가한다. 분석의 성능 저하를 막기 위해 우리는 여러 번의 분석을 통해 최적화된 길이로서 최대 경로 길이를 3으로 고정하였다.

PropFlow을 이용하여 얻을 수 있는 메타 경로는 그림 2과 같다. 그림 2의 실선은 반드시 있어야 하는 연결이며, 점선은 있을 수도 있는 연결이다. 그림 2를 참고하여 얻을 수 있는 기본 메타 경로는 $U - T$, $U-T-T$ 그리고 $U-U-T$ 이다.

메타 경로가 네트워크 스키마에서 검색되면, 표현 모델 G 의 각 사용자-주제 쌍(u, z)에 대해 토플러지 기능으로 사용하여 WPC (Weighted Path Count)를 적용한다. Path Count는 단순히 u 에서 시작하여 z 에서 끝나는 표현 모델의 메타 경로 인스턴스 수를 계산한다. 각종 경로 수는 각 메타 경로 인스턴스에 포함된 가장자리의 가중치를 합산한다. 사용자-주제 쌍과 이들에 대한 토플러지 특징이 얻어지면 로지스틱 회귀 분석을 진행한다. 회귀 분석으로 검증된 계수는 주어진 테스트 쌍 (u, z)에 토플러지 특징에 적용하여 주제에 대한 사용자 u 의 암시적 관심도를 예측한다.

4. 실험 및 평가

실험은 연구에서 제안한 암시적 관심사 분석 방법의 서비스 제공 속도와 정확도를 다른 암시적 관심사 분석 기법과 비교하여 검증한다. 특히, 지속적인 서비스 제공을 확인하기 위해 여러 번 제공 될 경우의 정확도를 확인한다.

우리는 공개적으로 사용 가능한 Twitter 데이터 세트 5개를 사용한다. 5개의 세트는 2017년 10~12월, 2018년 04월에 수집된 데이터[20]와 그리고 UCT에서 저장된 데이터 세트[21]를 이용한다. 이 데이터는 각각 다른 기간에 수집되었고, 약 천 만 개의 트윗이 포함되었다. 기존 연구에 따르면, 대부분의 트윗은 사용자 중 15%만이 2개월 동안 16개 이상의 트윗을 제공한다 [2]. 이 비율을 이용하여 우리는 사용자에 맞는 데이터를 적용하기 위하여 일부 트윗에 사용자를 임의로 부여하였다. 또한, 데이터의 확실성을 추가하기 위하여 사물의 웹 검색 데이터를 추가하였다. 웹 검색 데이터는 wikibase에서 제공된 archive 데이터를 이용한다[22].

우리는 실험을 위해 기존에 다른 연구에서 수집된 센서 정보를 날짜를 수정하여 사용하였다[14,15,23]. 특히, Reality[23]에서 진행된 연구에서 수집된 데이터에는 통화 기록, 근접한 Bluetooth 장치, 기지국 ID, 애플리케이션 사용 및 전화 상태(충전 및 유휴)가 포함되어있다. 추가로 상황과 위치정보는 수집된 트윗에서 관련 정보를 추출하여 임의 배정하여 상황 정보

로 사용하였다.

트윗의 텍스트에 주석을 달기 위해 권장 점수 임계값 0.1과 함께 TAGME RESTFUL API[12]를 사용하여 약 40만 개의 고유한 개념을 얻는다. 기존 연구에 따르면, TagMe의 선택은 트윗, 질의문 및 웹 페이지와 같은 다양한 유형의 텍스트에서 매우 합리적으로 작동한다고 이미 검증되었다. 우리는 데이터를 이용하여 주어진 시간 간격 T , 주제 Z 와 그리고 주제에 대한 각 사용자의 명시적 관심사 프로필을 추출한다. 프로필 추출을 위해 우리는 시간 간격에서 추출된 모든 개념을 단일 문서로 묶는다. 단일 문서 모음에 기본 규칙 변수 설정을 사용하여 LDA의 Gensim 구현을 적용한다[24]. 우리의 실험에서 시간 간격 T 의 길이는 1개월로 설정되었다. 또한, LDA에서 요구하는 주제 K 는 50으로 설정하였다. 이 값은 모든 주제 수에 대한 모든 실험을 반복하여 결정되었다.

주제 간의 관계를 학습하기 위해 우리는 요인별 항목-항목 협업 필터링 방법을 이용한다. 학습 단계에서는 Librec 라이브러리의 β , γ , λ , α 는 모두 0.001, 학습률은 0.01, 항목 잠재 요인 수는 10 그리고 반복 횟수는 500으로 설정한다.

실험에는 스마트폰이 스마트 공간에 진입한 순간부터 시작되어 데이터 교환까지 포함되어있다. 서비스 요청 후 제공까지 걸리는 서비스 지연 속도는 그림 3과 같다.



그림 3. 평균 서비스 제공 시간

기존 연구에 따르면, 메타 경로를 사용하는 암시적 관심사 분석 정확도는 비지도 링크 예측 PropFlow와 사용자 정의된 PathPredict 전략이 효율적이다[25]. 두 방법은 모두 높은 정확도를 가지는 것으로 검증되었다. 협업 필터링 방식보다 의미적 연관성 측정값이 더 효과적이다. 하지만, 우리는 상황에 맞는 서비스 제공을 위해 PropFlow와 하이브리드 필터링 방식을 혼합해 사용한다. 제안된 방법의 검증을 위해 기존 방법과의 비교 분석 실험을 한다. 우리는 지속 가능성을 고려하여 환경과 상황에 적합한 정보를 제공하는 것에 관심이 있다. 실험은 기존의

LISO에서 사용된 명시적 관심사 분석, PropFlow 분석, PathPredict 분석, 하이브리드 필터링 분석 그리고 우리가 제안한 방법을 비교한다.

사용자의 관심사를 탐지하는 것은 소셜 네트워크 데이터에서 리트윗 여부를 조사하는 방식으로 검증할 수 있다. 우리는 기존 연구에서 제안된 방법을 이용하여 리트윗 예측 애플리케이션에 미치는 영향을 평가하여 사용자 관심 감지 전략을 비교한다 [13]. 기존 연구에서는 다양한 작업이 제안되었지만, 우리는 사용자 관심사 기반의 알고리즘만 선택하여 검증한다. 검증 방법은 트윗을 평가하는 방법이므로, 우리는 상황 센서 정보를 텍스트 형태로 가공(예, 사용자 u가 2017년 10월 9일 레스토랑 A에 방문, 당근 케이크를 주문)하여 분석 데이터로 추가하였다.

알고리즘은 사용자 u에 대한 두 개의 연속 된 시간 간격의 트윗을 고려한다. 우리는 사용자가 2017년 10월에 게시하거나 리트윗 한 트윗을 기반으로 사용자의 관심 프로필 P(u)를 작성한다. 그리고 사용자가 리트윗 한 트윗을 리트윗 예측 적용 결과를 평가하기 위한 근거로 사용한다. 사용자 u의 리트윗을 예측하기 위해 2017년 11월과 12월 사이에 작성된 최소한 하나의 리트윗을 후보로 간주한다. 후보 트윗과 사용자 u의 사용자 관심 프로필 간의 주제 유사성을 다음과 같이 계산한다.

추출된 주제 $Z = \{z_1, z_2, \dots, z_k\}$ 에 대해 각 후보 트윗 d를 가중 벡터 $P(d) = (f_d(z_1), \dots, f_d(z_k))$ 로 나타낸다. $f_d(z_k)$ 의 값은 기존 연구에서 제공된 공식[13]을 수정하였으며, 공식 3과 같다.

$$f_d(Z_k) = \frac{\prod_{c \in D} w_{Z_k}(c)}{\sum_{i=1}^K \prod_{c \in D} w_{Z_i}(c)} \quad (3)$$

여기서 D는 트윗 d에 주석이 달린 개념 집합을 나타내고 $w_{Z_k}(c)$ 는 주제 $z \in Z$ 에서 개념 c의 확률이다. $P(d)$ 는 후보 트윗을 나타낸다. $P(u)$ 는 사용자 u의 관심 프로파일이다. $P(d)$ 와 $P(u)$ 의 코사인 유사성을 이용하여, 사용자 u의 트윗 d에 관심에 대한 유사성 점수를 계산한다. 그런 다음 유사성 점수를 기준으로 트윗의 순위를 내림차순 정렬한다. 후보 트윗의 순위 목록과 실제 트윗을 비교하여 리트윗 예측의 품질을 평가할 수 있다. 따라서 사용자의 관심도가 얼마나 확인되었는지를 검증할 수 있다.

평균 정밀도 (MAP)와 nDCG 랭크 k (nDCG @ k)를 두 가지 표준 정보 검색 메트릭으로 사용하여 예측 성능을 평가한다. 각 성능 비교는 표1과 같으며, MAP, nDCG @ 5, nDCG @ 10 및 nDCG @ 15로 실험이 진행되었다. 검색의 키워드에는 상황 정보나 위치 정보(ex, 레스토랑 A)가 사용되었다.

표 1. 리트윗 예측 애플리케이션의 성능 비교

방법	MAP	nDCG@5	nDCG@10	nDCG@15
		G@5	G@10	G@15
LISO	.024	.027	.031	.035
하이브리드 필터링	.026	.031	.032	.033
PathPredict	.033	.063	.058	.054
PropFlow	.034	.068	.062	.054
Our Approach	.034	.067	.061	.057

표 1의 결과를 기반으로, 우리가 제안한 접근 방식은 PropFlow와 비슷한 성능을 보인다.

결과적으로, 그림 3에서 나타난 것과 같이, 기존의 LISO에 비교해 서비스 제공 시간은 별다른 차이가 없는 것을 확인할 수 있다. 하지만, PropFlow와 PathPredict만을 이용한 암시적 관심사 분석은 제공 시간이 오래 걸려 모바일 서비스에 적합하지 않다. 그리고 하이브리드 필터링 방법을 사용하는 것은 제공 시간이 가장 적게 지연된다. 하지만, 표 1의 결과에서 하이브리드 방법만을 사용할 경우 암시적 관심사 분석 방법의 정확도가 다른 방법에 비해 낮다. 그러므로, 본 연구에서 제안한 방법이 스마트 공간에서는 효과적임을 알 수 있다.

III. 결 론

우리는 스마트 공간을 고려한 사용자 관심사가 적용된 지능형 프레임워크를 제안하였다. 기존에 사물에 암시적 분석을 시도한 연구는 존재하지 않는다. 제안된 알고리즘은 명시적 관심사를 제공하는 기존 연구와 비교하면 성능을 크게 벗어나지 않으면서, 암시적인 관심사를 고려한 정보를 사물에서 서비스로 제공할 수 있다. 특히, 암시적 관심사는 다른 분야에서 정보가 부족한 콜드 스타트나 정보 누락의 문제 해결에 좋은 결과를 얻은 연구가 존재한다. 스마트 공간에서도 적절한 대응이 가능한지에 대한 추가 연구가 필요하다. 마지막으로, 본 논문에서는 스마트 공간에서 개인 정보 보안 문제점은 고려하지 않았다. 개인 정보 보안 문제는 지속해서 발전 가능성이 있는 스마트 공간에서는 중요 이슈 중 하나이며, 향후 관련 내용에 관한 추가 연구가 필요하다.

REFERENCES

- [1] Henniger Olaf, Naser Damer, Andreas Braun, "Opportunities for biometric technologies in smart environments," *Proceedings of 13th European Conference on Ambient Intelligence*, pp. 175–182, Malaga, Spain, Apr. 2017.
- [2] Fabian Abel, Qi Gao, Geert-Jan Houben, Ke Tao, "Analyzing user modeling on twitter for personalized news recommendations," *19th International conference on user modeling, adaptation, and personalization*, pp. 1–12, Berlin, Heidelberg, Jul. 2011.
- [3] Y. Li, J. Jiang, T. Liu, M. Qiu, X. Sun, "Personalized microtopic recommendation on microblogs," *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 8, Issue 6, No. 77, pp 1–21, Aug. 2017.
- [4] 김정인, 박상진, 김형주, 최준호, 김한일, 김판구, "나이브 베이즈 기반 소셜 미디어 상의 신조어 감성 판별 기법," *스마트미디어저널*, 제9권, 제1호, 51–59쪽, 2020년 3월
- [5] 홍택은, 김정인, 신주현, "인스타그램 이미지와 텍스트 분석을 통한 사용자 감정 분류," *스마트미디어저널*, 제5권, 제1호, 61–68쪽, 2016년 3월
- [6] 장서윤, 강지훈, "위치 정보 공유 사물 (LISO): 사용자 관심을 기반으로 소셜 미디어 데이터를 사용하는 상황 인식 지능형 서비스," *정보과학회논문지*, 제47권, 제9호, 814–819쪽, 2020년 9월
- [7] Nemanja Spasojevic, Jinyun Yan, Adithya Rao, Prantik Bhattacharyya, "Lasta: Large scale topic assignment on multiple social networks," *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1809–1818, New York, USA, Aug. 2014.
- [8] Nacéra Bennacer Seghouani, Coriane Nana Jipmo, Gianluca Quercini, "Determining the interests of social media users: two approaches," *Information Retrieval Journal*, Vol. 22, No. 1–2, pp. 129–158, Jul. 2019.
- [9] 임명진, 김판구, 신주현, "리뷰의 의미적 토픽 분류를 적용한 감성 분석 모델," *스마트미디어저널*, 제9권, 제2호, 69–77쪽, 2020년 6월
- [10] Silvia Calegari, Gabriella Pasi, "Personal ontologies: Generation of user profiles based on the YAGO ontology," *Information processing & management*, Vol. 49, No. 3, pp. 640–658, May 2013.
- [11] Pavan Kapanipathi, Prateek Jain, Chitra Venkataramani, Amit Sheth, "User interests identification on twitter using a hierarchical knowledge base," *Proceedings of 11th European Semantic Web Conference*, pp. 99–113, Anissaras, Greece, May 2014.
- [12] Ferragina Paolo, and Ugo Scaiella, "Tagme: on-the-fly annotation of short text fragments (by wikipedia entities)," *Proceedings of the 19th ACM international conference on Information and knowledge management*, pp. 1625–1628, Toronto, Canada, Oct. 2010.
- [13] J. Wang, WX. Zhao, Y. He and X. Li, "Infer user interests via link structure regularization," *ACM Transactions on Intelligent Systems and Technology (TIST)* Vol. 5, No. 23, pp. 1–22, Apr. 2014.
- [14] Alberto Rosi, Marco Mamei, Franco Zambonelli, "Integrating social sensors and pervasive services: approaches and perspectives," *International Journal of Pervasive Computing and Communications*, Vol. 9, No. 4, pp.294–310, Nov. 2013.
- [15] Z. Mao, Y. Jiang, G. Min, S. Leng, X. Jin, and K. Yang, "Mobile social networks: Design requirements, architecture, and state-of-the-art technology," *Computer Communications*, Vol. 100, pp. 1–19, Mar. 2017.
- [16] Federico Barrios, Federico López, Luis Argerich, Rosa Wachenchauzer, "Variations of the similarity function of textrank for automated summarization," *arXiv preprint*, arXiv:1602.03606, Feb. 2016.
- [17] 장서윤, 박종현, "wikidata를 이용하는 상황 인지 엔티티 링크 프레임워크," *한국컴퓨터정보학회 학술발표논문집*, 제28권, 제2호, 587–589쪽, 2020년
- [18] Yong Wang, Jiangzhou Deng, Jerry Gao, and Pu Zhang, "A hybrid user similarity model for collaborative filtering," *Information Sciences*, Vol. 418–409, pp. 102–118, Dec. 2017.
- [19] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, Philip S. Yu, "A survey of heterogeneous information network analysis," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No. 1, pp. 17–37, Aug. 2016.
- [20] Internet archive(2018).
<https://archive.org/details/archiveteam-twitter-stream-2018-04>,
<https://archive.org/details/archiveteam-twitter-stream-2017-10> (accessed Feb., 21, 2021).
- [21] UCT DataSet(2015).

<https://bitbucket.org/sliang1/uct-dataset/get/UCT-Dataset.zip> (accessed Feb., 21, 2021).

[22] wikibase data Set(2018).

<https://archive.org/details/wikibase-wikidatawiki-20180120> (accessed Feb., 21, 2021)

[23] Nathan Eagle, Alex (Sandy) Pentland, "Reality mining: sensing complex social systems," *Personal and ubiquitous computing*, Vol. 10, No. 4, pp. 255-268, Nov. 2006.

[24] GENSIM(2020),

<https://radimrehurek.com/gensim/models/ldamodel.html> (accessed Feb., 21, 2021).

[25] Fattane Zarrinkalam, Mohsen Kahani, Ebrahim Bagheri, "Mining user interests over active topics on social networks," *Information Processing & Management*, Vol. 54, No. 2, pp. 339-357, Mar. 2018.

저자 소개



장서윤(정회원)

2009년 충남대학교 컴퓨터공학과 학사.

2016년 충남대학교 컴퓨터공학과 네이터 및 소프트웨어 전공 석박통합과정 수료.

~ 현재 충남대학교 컴퓨터융합학부 연구원

<주관심분야 : 상황인지, XML, 웹정보처리 및 분석, 관심사 분석>



강지훈(정회원)

1979년 서울대학교 계산통계학 학사.

1981년 한국과학기술원 전한석 석사.

1996년 한국과학기술원 전산학 박사.

~ 현재 충남대학교 컴퓨터융합학부 교수

<주관심분야 : 상황인지, XML, 웹정보처리 및 분석>