

T-Commerce를 위한 위치인식 및 텍스트마이닝 기반 사용자 구매 패턴 연구

*송혜진 김진아 이선민 **문남미

호서대학교 컴퓨터학과 ** 컴퓨터소프트웨어전공

*song120408@naver.com

A Study on User's Purchasing Pattern based on Text mining and Location awareness for T-Commerce

*HyeJin Song Jin-Ah Kim Sunmin Lee ***Nammee Moon

Hoseo University

요약

최근 TV시청은 다양한 매체를 통해서 이루어지고 있으며, 특히 스마트폰을 통한 시청률이 늘고 있는 상황이다. 광고시장에서도 TV시청 중에 스마트기기를 함께 이용하는 멀티태스킹 사용자가 급증하고 있으며 특히 10~30대의 사용이 적극적이다. TV시청 중 스마트 기기의 사용분야는 메신저, 정보검색, SNS 순이며 스마트 기기사용 내용 중 69%는 시청하던 TV 시청과 관련된 것이었다. 이 중에 75%는 TV에 등장한 제품, 브랜드, 장소에 관한 것이다[1]. TV를 시청하는 상황에 스마트기기의 소셜 활동의 문자를 분석하는 것은 사용자 의도를 파악할 수 있는 의미가 있으며, 시청자의 현재 위치를 파악함으로써 시청자의 의도에 반영되는 상황을 파악할 수 있다. T-Commerce 구매 의도는 사용자의 현재 상황에 대한 순간 의도를 파악하는 것이 중요하며, 이와 같은 구매의도를 파악하기 위해서 본 연구에서는 GPS와, Wi-Fi 기반 Fingerprinting 측위기법을 사용하여 특별한 도구나 장비의 설치 없이 현재위치와 멀티태스킹 데이터를 분석하여 구매의도를 파악한다. T-Commerce 소비환경 패턴이 바뀔에 따라, 다양한 소비 환경 데이터 분석은 효율적인 광고 제공과 만족도를 높일 것으로 기대된다.

1. 서론

T-commerce의 정의는 TV Based e-Commerce이다. 즉 TV의 사용편의성과 인터넷의 즉시성이 결합되어진 개념이다.[2]. 2000년 초반기에 T-Commerce 가치사슬과 현재 가치사슬은 많은 차이를 갖고 있으나 사용자의 다양한 욕구를 반영하기 위한 기본 개념에는 변화가 없다. 가장 큰 변화는 TV와 스마트폰, 기타 인터넷 활용은 더 이상 분리되어진 대상이 아닌 상호간에 시너지를 일으킬 수 있는 관계가 형성되어 있다는 것이다[1].

최근 방송소비자행태조사에 따르면, TV를 시청할 때 스마트폰을 사용하는 이들의 69%는 스마트폰으로 TV관련 소통과 검색을 하고 있으며, 그 중 75%는 TV속 제품에 관해서 알아보고 있는 것으로 조사되었다.[1] 또한 멀티태스킹은 TV만 시청하는 이들보다 시청한 프로그램의 광고를 1.5배 더 많이 기억하는 것으로 나타나고 있기 때문에 PPL을 직접 검색하거나 구매한 경험이 높은 비율로 나타나고 있다.

TV의 모형이나 기기는 꾸준히 변화되어 왔으나, 지난 50년 동안 집안의 엔터테인먼트와 정보를 제공하는 주 도구로서의 역할은 변화하지 않았다. 가장 큰 변화는 TV와의 상호작용이 어느 때보다 중요하다는 것이다. LAN 혹은 Wi-Fi 브로드밴드와의 연결, 미디어 공유, 스마트폰과의 연결(2nd Screen)등 기기의 변화와 더불어 사용자의 요구가 바뀌었기 때문에 보다 사용자의 경험(UX: User Experience)에 맞춰진 인터페이스와 콘텐츠의 제공이 필요한 시점이다[3].

본 연구는 사용자의 의도와 경험을 파악하기 위하여, 사용자 위치 데이터와 멀티태스킹기반 소셜 활동데이터를 추적하고 분석하여, 기존 T-Commerce 데이터와 융합시켜 활용하는 것을 목표로 한다.

2. 시스템 개요



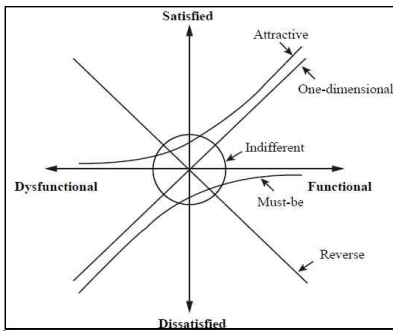
(그림 1) 시스템 개요

본 연구는 영상 내에 타이틀, 서브타이틀, 상영시간, 자막정보등과 스마트 기기를 이용한 SNS 관련 정보 그리고 Wi-Fi AP와 GPS를 이용하여 실내·외 환경에서의 위치를 파악한 후 수집되어진 정보를 분석

[4],[9]하여 T-Commerce 사용의도를 분석 및 분류하고, 향후 소비패턴을 예측 분석하고자 한다. (그림 1)은 시스템 개요를 설명한다.

2. T-commerce 소비 패턴 분석

KANO 모델은 제품이나 서비스의 품질 속성에 대한 고객의 주관적 만족도를 효과적으로 측정할 수 있는 도구이다. 양방향으로 제공하는 다양한 TV를 기반으로 한 상거래에 공통요인으로 접근성, 신뢰성, 신속성, 안정성, 정보성이 만족과 불만족에 큰 영향을 미치는 것으로 연구되어졌다[5]. 방송과 통신환경이 경계를 허물며 SNS등을 통하여 방송에 소비자가 실시간으로 직접참여 가능해 지고 있으며 고객과의 상호 작용이 용이해지고 이에 따라 생방송 중에도 고객이 방송에 참여하는 반응성은 높은 매력적 요인으로 조사되고 있다. 또한 즉시적 영향을 나타내지 않아도 고객의 관심은 향후 구매로 이어질 수 있기 때문에 소비자의 행동패턴을 분석하는 것은 의의가 있다.



(그림 2) Kano 2차원 소비만족도모델

3. SNS 기반 데이터분석

방송 프로그램의 시청률과 SNS를 포함하는 온라인 데이터는 상관관계가 있다[1][6]. 오피니언마이닝 연구는 온라인 뉴스와 소셜 미디어의 코멘트, T-Commerce의 상품평등 사용자가 다양한 콘텐츠를 통해 표출한 의견을 추출, 분류, 이해, 자산화 하는 과정을 의미한다. 감성분석(Sentimental Analysis)의 다양한 기술을 활용하여 수행되어지며, 문서에 포함된 상호간 의견 정보와 블로그, 카페, 게시판 등 온라인 상에 새로운 상품이나 서비스에 나타내는 의견을 분석하여 반응을 반영하는 것이다.

예) 핸드폰 블루투스 이어폰
 사용자의견 : 음질은 좋지만, 디자인이 투박하고 불편하다.
 (+1) (-1) (-1)
 평판 : +1-1-1= -1 : 소비자 반응은 부정이라고 평가

본 연구에서는 이를 위해 선행적인 연구로 텍스트마이닝 기법을 사용하였다. 텍스트마이닝은 대용량의 문서로부터 유용한 패턴을 찾기 위한 목적으로 텍스트 자료에 통계와 기계 학습 분야의 알고리즘과 방법을 적용하는 것을 의미한다. 키워드를 추출하기 위하여, 형태소 분석을 실시하여 명사에 해당하는 키워드를 추출한다. 추출된 명사는 때에 따라 여러 문서 혹은 모든 문서에서 공통적으로 나타나서 별다른 의미를 주지 못하는 경우가 존재하므로, 불용어 사전을 정의해서 이를 제거

할 수 있도록 한다. 비정형 자료인 문서를 정형화하기 위해서, 단어-문서 행렬을 만들고 i개의 단어와 j개의 문서로 이루어진 행렬의 형태로 나타내 빈도수를 측정할 수 있도록 한다. 또한 단어와 문서 상호관계를 유기적으로 표현하기 위하여, 가중치 함수를 사용한다. 가중치 함수 (W_{ij})는 지역가중치($L(i, j)$)와 전역가중치($G(i)$)를 사용하여 곱으로 계산해서 활용한다.

$$W_{ij} = L(i, j) \times G(i)$$

<표1> 단어-문서 행렬

* f_{ij} 는 i 단어가 j 문서에 포함된 빈도수를 의미한다.

	Doc 1	...	Doc j	...	Doc n
Term 1	f_{11}	...	f_{1j}	...	f_{1n}
:	:	...	:	...	:
Term i	f_{i1}	...	f_{ij}	...	f_{in}
:	:	...	:	...	:
Term m	f_{m1}	...	f_{mj}	...	f_{mn}

4. 위치 데이터 수집 및 분석

실외환경과 실내 환경 두 곳에서 이루어지는 위치 데이터 수집을 위해 각각의 베이지 규칙을 적용한 확률모델기반 Naive Bayes 분류기를 사용[8]하여, 실내·외 환경을 구분 지은 후, 실외환경에서는 GPS를 실내 환경에서는 Wi-Fi기반 Fingerprinting 기술을 사용한다.

GPS를 이용한 실외 측위 기법은 Google Map을 기반으로 하여, 현재 장소의 GPS 좌표 값과 일정한 반경 범위 안에 있는 상점과 같은 건물의 GPS 좌표를 토대로 현재 장소를 등록한다. 이때, GPS좌표 외에 시간, 요일, 이동상태도 함께 저장한다.

실내 측위 기법은 Wi-Fi 신호로 라디오 맵을 구축하는 단계와 구축된 정보를 바탕으로 위치를 비교해 추정하는 측위 단계로 나누어진 다[4],[7]. Wi-Fi 라디오 맵을 구축하기 위해서는 위치 측위 하고자 하는 공간을 격자로 구분하여 각 구분된 곳 마다 주변 AP로 부터의 신호 정보를 DB로 만들고, 구축된 맵을 바탕으로 실제 Wi-Fi 신호로 측위된 위치와 비교하여 가장 유사한 위치를 선택해 추정이 이루어지게 된다. 일반적으로 넓은 공간에서는 AP를 여러 개 사용하여야 하나, 50평 이내의 실내 거주공간에서는 적은 수의 AP를 사용해 위치를 추정해도 정확도를 손상시키지 않는다.

<표 2> 위치 데이터 구조 예시

PCode	Location	Pname	G	UID	Etime	Ltime	STime	Date
1101	실외	상점1	1	141156	16:30	16:37	7	16-09-07
1102	실외	상점2	1	141156	16:34	17:24	50	16-09-07
1201	실내	침실1	0	154132	11:24	11:37	13	16-09-08
:	:	:	:	:	:	:	:	:
1501	실내	침실2	0	124856	18:37	18:55	18	16-09-10
1502	실내	화장실	0	147543	10:14	10:30	16	16-09-11
1503	실외	식당1	1	147543	10:27	10:47	20	16-09-11

데이터 수집단계를 거쳐 축적된 데이터는 <표2>의 예시와 같은 데이터를 갖는다. 장소와 날짜, 사용자 ID, 시청시간들이 기록과 함께, 같은 공간에 사용자 외에 다른 사용자가 있는지를 파악해서 기록한다. T-Commerce 특성상 상호작용이 중요한데 본인 이외에 다른 사용자

가 있어서 행위에 변화요인을 제공할 수 있기 때문에 이를 반영한 것이다.

5. 결론

T-Commerce 구매 의도는 사용자의 현재 상황에 대한 순간 의도를 파악하는 것이 중요하다. 이를 위하여 사용자의 SNS 기반 텍스트 마이닝을 수행하며, 구매의도를 파악하기 위해 본 연구에서는 GPS 기술과 Wi-Fi 기반 Fingerprinting 측위기법을 혼용하여 특별한 도구나 장비의 설치 없이 현재위치와 멀티태스킹 데이터를 분석하여 구매의도를 파악한다. T-Commerce 소비환경패턴이 바뀔에 따라, 다양한 소비 환경 데이터 분석을 효율적으로 제공함으로써 효과적인 광고 제공과 더불어 소비자 만족도를 높일 것으로 기대된다.

참고문헌

- [1] 2016 소비자행태조사, 한국방송광고진흥공사, 2016
- [2] 문남미 "T-commerce 비즈니스 전략과 기술", 방송공학회지 제 8권 특집 최신디지털방송기술 II, pp209-222, 2003
- [3] Andrea Ingrosso 외 4, "UX and Usability on Smart Tv : Case Study on A t-Commerce Application", Vol 9191 of the Series LNCS pp.312-323, 2015.7
- [4] 김진아 외 1 "실내 공간 활용증대를 위한 Wi-Fi 기반 위치 인식 시스템 설계", 정보처리학회 2016 추계학술대회논문지
- [5] 박준용 외 1, "TV기반 상거래의 품질 속성분석을 통한 소비자 만족도 증대요인 분석", 한국전자거래학회지 제 21권 제2호, pp 61-80, 2016
- [6] 마경란, "SNS 빅데이터 분석을 통한 텔레비전 프로그램 시청률에 대한 연구", 한양대학교 석사논문 2013.08
- [7] 김광열, 신요안, "실내 Wi-Fi Fingerprinting 무선측위 기법의 Virtual Radio Map 구축 특성 분석", 한국통신학회 학술대회 논문집, pp.269-270, 2015.1
- [8] 서호 외 1명, "스마트폰 로그마이닝을 통한 사용자 의미장소의 자동인식", 한국컴퓨터종합학술대회 논문집, 2014.06
- [9] 최승환 외 4명, "다중 GPS를 이용한 EKF 기반의 실외 위치 추정 시스템", 로봇학회 논문지 8(2), 2013.06