
마이크로 블로깅 서비스를 지원하기 위한 컨텍스트 모델 기반 자동 블로깅 시스템

박재민*, 이상용**

An Auto-blogging System based Context Model for Micro-blogging Service

Jae-min Park*, Sang-yong Lee**

요약 소셜 네트워크 서비스의 가장 대표적인 마이크로 블로깅 서비스를 효과적으로 제공하기 위해 사용자가 자신의 현재 상황정보를 간편하게 기록하고 그 정보를 바탕으로 다른 사람들과 네트워크를 형성하고 유지하도록 하는 것이 중요하다. 하지만 모바일 환경에서 사용자가 자신의 정보를 매번 모바일 디바이스를 통해 직접 입력하는 것은 매우 번거로운 작업이다. 본 논문에서는 획득된 사용자 컨텍스트를 이용하여 사용자의 현재 행동과 다음 목적지를 추론한 후, 자동으로 문장을 생성하여 블로깅을 해주는 컨텍스트 모델 기반 자동 블로깅 시스템을 제안한다. 컨텍스트 모델을 생성하기 위해 사용자의 행동 추론은 나이브 베이즈 분류기를 이용하고, 이동중인 사용자의 다음 목적지 추론은 시퀀스 매칭을 이용하였다. 생성된 컨텍스트 모델을 기반으로 5WIH 구조를 이용하여 상황에 적합한 문장을 생성하여 자동으로 블로깅하였다. 제안한 방법의 정확도를 평가한 결과 평균 88.73%의 정확도를 보여 자동 블로깅 서비스 가 가능성을 보여주었다.

주제어 : 소셜 네트워크 서비스, 컨텍스트, 나이브 베이즈 분류기, 시퀀스 매칭

Abstract Social network service is service that enables the human network to be built up on web. It is important to record users' information simply and establish the network with people based on the information to provide with the social network service effectively. But it is very troublesome work for the user to input his or her own information on the mobile environment. In this paper we suggested a system which classifies users' behavior using context and creates blogging sentences automatically after inferring the destination. For this, users' behavior is classified and the destination is inferred with the sequence matching method using Naive Bayes classification. Then sentences which are suitable for situation is created by arranging the processed context using the structure of 5WIH. The system was evaluated satisfaction degree by comparing the created sentences based on actually collected data with users' intension and got accuracy rate of 88.73%.

Key Words : Social Network Service, Context, Naive Bayes Classifier, Sequence Matching

1. 서론

개인간의 연결을 통해 인맥을 형성하는 서비스는 과거 향우회나 동창회, 종친회 등의 형태로 존재해왔다. 이러한 인맥 관계를 온라인 상으로 도입하여 지인과의 인맥 관계를 형성할 수 있도록 한 것이 소셜 네트워크 서

비스이다. 최근 트위터의 폭발적 인기의 영향으로 모바일 환경에서 단편적인 감상 등을 블로깅하는 마이크로 블로깅 형태가 주를 이루게 되었다. 하지만 모바일 환경에서 이동중인 사용자가 자신의 정보를 매번 모바일 디바이스를 통해 직접 입력하는 것은 매우 번거로운 작업이다. 소셜 네트워크 서비스를 효과적으로 제공하기 위해서는 간

*공주대학교 컴퓨터공학과(soulsieg@kongju.ac.kr)

**공주대학교 컴퓨터공학부(sylee@kongju.ac.kr), 교신저자

논문접수: 2012년 5월 1일, 1차 수정을 거쳐, 심사완료: 2012년 5월 14일

편하게 사용자의 정보를 기록하고, 그 정보를 공유함으로써 다른 사람들과 네트워크를 형성하고 유지할 수 있도록 하는 것이 중요하다.

본 논문에서는 사용자의 상황을 문장으로 생성하고 자동으로 블로그 해주는 컨텍스트 모델 기반 자동 블로그 시스템을 제안한다. 시스템은 모바일 환경에 뛰어난 나이브 베이스 분류기를 이용하여 사용자의 행동을 추론하고, 사용자의 이동 경로 등 반복적으로 발생하는 상황에 대해 높은 정확도를 보이는 시퀀스 매칭 기법을 적용하여 사용자의 다음 목적지를 추론한다. 또한, 수집된 컨텍스트들을 이용하여 사용자의 상황을 문장으로 생성하기 위해 4W(Who, When, Where, What) 기반 컨텍스트 모델을 사용하였다.

2. 관련 연구

2.1 소셜 네트워크 서비스

오늘날 웹 이용자들은 웹을 단순한 검색 수단이 아닌 자신의 의사를 표현하고 정보를 공유하며 상호교류를 하기 위한 커뮤니케이션 공간으로 활용하고 있다. 이러한 변화를 주도하고 있는 것이 소셜 네트워크 서비스(Social Network Services; SNS)이다. 방송통신위원회와 한국인터넷진흥원이 발표한 2010년 인터넷이용실태조사 보고서에 따르면 인터넷 이용자의 65.7%가 SNS를 이용하고 있으며 10대에서 30대는 대다수, 40대도 50%에 이르는 이용률을 보이고 있다[1]. 대표적인 SNS로는 마이스페이스, 페이스북, 싸이월드, 트위터, 미투데이 등이 있으며 많은 사용자들이 사용하고 있다. 각 SNS는 나 자신의 모든 것을 기록하면서 그 데이터를 바탕으로 다른 사람들과 네트워크를 형성하는 유사한 형태를 가지고 있다. 기존의 소셜 네트워크 서비스는 획득한 사용자 정보를 가공하여 정보를 분류하고 저장하는 방법에 대해 연구되고 있지만 자동으로 블로그를 제공하는 서비스는 제공하지 못하고 있다[2].

2.2 컨텍스트 기반 데이터 모델링

컨텍스트 기반의 데이터 모델링은 다양한 단말기와 응용서비스를 통해서 데이터를 분류하고 가공하여 상황에 맞는 서비스를 제공하기 위한 기술이다[3]. 컨텍스트 모델링을 위한 방법으로는 Bayesian Networks, Markov

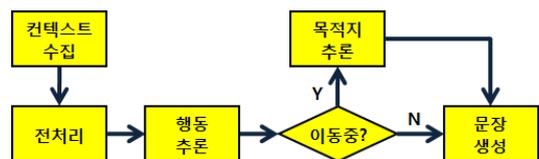
Model, Sequence Matching 등의 방법이 사용되고 있다. 이 중 나이브 베이스 분류기는 베이스 정리를 기초로 한 것으로, 단순한 학습 방법에도 불구하고 다른 분류방법들보다 뛰어난 추정능력을 발휘하여 모바일 환경에 적합하다고 알려져 있다[4].

시퀀스 매칭 기법은 저장된 데이터 시퀀스 중 사용자에게 주어지는 질의 시퀀스와 유사한 데이터 시퀀스를 검색하는 방법으로 유사 시퀀스 매칭이라고도 한다[5]. 예를 들어, 사용자의 현재 위치를 포함한 최근의 위치 정보들을 하나의 집합으로 하여 질의 시퀀스를 생성하게 될 경우 현재까지 저장된 위치 정보들은 데이터 시퀀스로 하여 시퀀스 매칭을 시행하게 된다. 시퀀스 매칭 기법은 연속 속성을 갖는 컨텍스트를 다루기에 적합하고 사용자의 이동 경로 등 반복적으로 발생하는 상황에 대해 높은 정확도를 보인다[6].

3. 자동 블로그 시스템

본 논문에서는 모바일 환경에서 사용자의 행동을 추론하는데 적합한 나이브 베이스 분류기와 위치 추론에 뛰어난 시퀀스 매칭을 이용하여 컨텍스트 모델을 생성하고 자동 블로그를 제공하는 시스템을 제안하였다.

시스템의 구조와 처리 흐름은 그림 1과 같다. 소셜 네트워크 서비스 이용 요청이 들어오면 각 센서를 통해 얻어진 사용자의 컨텍스트를 전처리하여 행동 추론기에 보내 사용자의 행동을 추론하게 된다. 추론된 결과가 현재 사용자가 이동 중일 경우 다음 목적지를 추론하게 되고, 이동 중이 아니라면 문장 생성기를 이용하여 사용자의 현재 행동에 관련된 문장을 소셜 네트워크 서비스에 블로그한다.



[그림 1] 시스템의 구조와 처리 흐름

예를 들어, 한 사용자가 학교에서 공부를 하던 중 자신의 현재 위치와 행동을 지인들에게 알리고자 시스템에 블로그 요청을 하였다면, 시스템은 현재 사용자의 주변

정보를 센싱하여 전처리를 하게 된다. 행동 추론기는 전처리를 마친 데이터와 지금까지 축적된 데이터와 비교하여 현재 사용자가 공부를 하고 있다는 것을 파악하게 된다. 추론된 행동 정보는 문장 생성기를 통해 ‘아침에 나는 학교에서 공부를 하고 있다’라는 문장으로 생성하여 소셜 네트워크 서비스에 자동으로 블로그 한다.

3.1 컨텍스트 수집

컨텍스트 수집 과정에서는 모바일 기기의 각 센서를 통하여 사용자의 컨텍스트 데이터를 지속적으로 수집한다. 수집 방법은 일정 범위를 이동하였을 때 수집하는 방법과 미리 입력된 장소에 접근하였을 때 수집하는 두 가지 방법을 기본으로 한다. 그리고 건물 안과 같이 GPS 수신에 제한된 장소에서는 사용자가 수동으로 수집할 수 있도록 하였다. 수집되는 컨텍스트는 표 1과 같으며, 레코드 셋 기반으로 데이터 베이스에 저장된다.

〈표 1〉 컨텍스트의 종류와 형태

구분	컨텍스트	데이터 형태
GPS	위도	36.473762
	경도	127.150247
가속도	동작 여부	475
날씨	날씨	맑음, 흐림
SYSTEM	날짜	2011-10-30
	요일	Monday
	시간	13:20:30
어플리케이션	어플리케이션	메모, 게임, ..., 버스시간

3.2 전처리

전처리 과정에서는 수집된 컨텍스트를 실생활에 사용되는 언어로 추상화시킴으로써 소셜 네트워크 서비스에 블로그된 문장이 실제 사람이 작성한 것처럼 느껴지도록 한다. 표 2는 수집된 시간을 아침, 점심, 저녁 시간 등으로 매치시켜 실제 블로그에 이용될 수 있도록 해주는 시간 구분표이다.

〈표 2〉 시간 관련 컨텍스트

시간	구분	시간	구분
0시	밤	12시~13시	점심
1시~5시	새벽	14시~17시	오후
6시~9시	아침	18시~21시	저녁
10시~11시	오전	22시~23시	밤

장소의 경우 위도와 경도를 그대로 사용할 수 없기 때문에 표 3처럼 해당 위치와 장소명을 매핑하여 사용한다.

〈표 3〉 장소 관련 컨텍스트

위도	경도	장소명
36.852436	127.150247	8공학관
36.851975	127.150269	연구실
:	:	:

3.3 행동 추론

행동 추론 과정에서는 컨텍스트 정보들을 이용하여 사용자의 상황을 정확하게 판단한다. 이 과정에서 Statistics Canada에서 조사한 GSS를 기반으로 7가지의 대분류와 16가지의 소분류로 사용자의 행동을 분류하여 표 4와 같이 정리하였다[7].

〈표 4〉 사용자 행동 분류표

대분류	소분류
업무	업무, 회의
교육	수업, 세미나
사회 활동	만남, 식사
문화 활동	영화, 공연, 봉사활동
개인 활동	종교, 휴식, 수면
활동적 취미	운동, 음악
비 활동적 취미	독서, 컴퓨터

나이브 베이즈 분류기는 수집된 사용자의 컨텍스트에 분류된 행동을 추가한 행동패턴 테이블을 학습 데이터로 가지게 된다. 새 데이터 F가 입력될 경우 행동패턴 테이블을 C로 정의하면 식 (1)과 같이 새 데이터가 각 분류에 속할 확률을 계산하게 된다.

$$\frac{P(C\text{대분류})}{P(C)} \times \frac{P(C,F\text{.day})}{P(C\text{대분류})} \times \frac{P(C,F\text{.week})}{P(C\text{대분류})} \times \dots \times \frac{P(C,F\text{.apps})}{P(C\text{대분류})} \quad (1)$$

3.4 목적지 추론

목적지 추론 과정은 행동 추론 결과가 ‘이동 중’으로 나왔을 때 컨텍스트 DB에서 얻은 위치 정보를 일별로 분류하여 데이터 시퀀스를 생성하고, 입력된 위치의 변동을 절의 시퀀스로 하여 사용자의 목적지를 추론하게 된다.

〈표 5〉 목적지 추론

sequence	location				frequency
시퀀스 a	1	2	4	3	15
시퀀스 b	1	2	4	2	11
시퀀스 c	1	2	4	5	10
시퀀스 d	1	3	4	2	12
질의 시퀀스	1	2	4		

표 5는 질의 시퀀스 '1, 2, 4'가 입력되었을 때의 예로 목적지 추론기는 데이터 시퀀스에서 질의 시퀀스와 일치하거나 유사한 시퀀스를 찾게 된다. 이때 질의 시퀀스와 일치하는 시퀀스 a, b, c 중 가장 빈도수가 높은 시퀀스 a의 다음 위치인 3을 목적지로 추론하게 된다. 목적지를 추론한 정확도는 데이터 시퀀스의 축적량에 비례하여 증가되는 것을 확인하였으며 최종 평균 90% 이상의 정확도를 보이는 것으로 평가되었다.

3.5 문장 생성

문장 생성 과정에서는 행동 추론기에서 추론된 결과를 바탕으로 4W의 컨텍스트를 표 6과 같은 형태의 문장으로 생성하게 된다.

〈표 6〉 문장 생성

분류	4W	컨텍스트 공급자	컨텍스트 내용	기본 문장
행동	Who	SYSTEM	등록된 사용자	[When]에 [Who]는 [Where]에서 [What]을 하고 있다.
	When	컨텍스트 DB	추상화된 시간	
	Where	컨텍스트 DB	현재 장소	
	What	행동 추론기	추론된 행동	
이동중	Who	SYSTEM	등록된 사용자	[When]에 [Who]는 [Where]로 향하고 있다.
	When	컨텍스트 DB	추상화된 시간	
	Where	목적지 추론기	추론된 위치	

문장은 행동과 이동 중의 2가지로 분류하여 제공한다. 예를 들어, Who에는 사용자 자신, When에는 아침, Where에는 도서관, What에는 공부라는 데이터가 들어왔을 경우 '아침에 나는 도서관에서 공부를 하고 있다.'라는 문장이 생성된다. 사용자가 이동 중이면서 Who에는 사용자 자신, When에는 점심, Where에는 식당이 들어올 경우 '점심에 나는 식당으로 향하고 있다.'라는 문장이 생성된다.

4. 실험 및 평가

4.1 설계 및 구현

실험을 위한 장비 사양과 환경은 표 7과 같다.

〈표 7〉 실험환경

장비	사양
Mobile Device	Motorola 'Defy' HTC 'Nexus One' LG 'Optimus Z'
CPU	TI OMAP3610 800MHz QSD 8250 1GHz QSD 8650 1GHz
RAM	512MB
OS	GingerBread
Language	Java for Android, XML
DB	SQLite3
Programming Tool	Eclipse

그림 2는 컨텍스트 수집기의 화면으로, 왼쪽은 센싱 화면으로 사용자가 수동으로 컨텍스트를 수집하는 수집 버튼과 자주 가는 장소를 입력할 수 있는 위치 입력 버튼, GPS가 작동하지 않는 건물 안에서 위치를 선택해 수집하게 되는 수동 입력 버튼이 제공된다. 오른쪽은 장소를 등록하는 위치 입력 버튼으로 장소명을 입력하고 확인을 누르게 되면 해당 위도, 경도에 매치되는 장소가 저장된다.



[그림 2] 컨텍스트 수집기

그림 3은 행동 추론 결과 화면으로, 새로 입력된 데이터는 EditText에서 확인할 수 있어 사용자가 수동으로 데이터를 입력하거나 수정할 수 있다. 추론이 시작되면 행동 추론 모듈은 입력 데이터가 1번부터 7번까지의 행동 대분류에 속할 확률을 구하게 되고 가장 높은 확률의 대분류를 기준으로 소분류의 확률을 구한다. 추론된 대분류와 소분류의 행동 정보는 입력 데이터와 함께 DB로 저장되어 이후 새로운 입력 데이터를 추론할 때 정확도를 높게 된다.



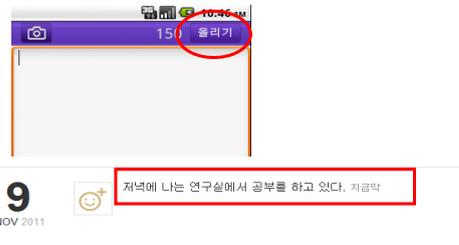
[그림 3] 행동 추론

그림 4는 시퀀스 매칭 기반의 목적지 추론 화면으로, 사용자의 현재 위치로부터 DB에 저장된 이전 위치를 기반으로 질의 시퀀스를 생성한다. 추론을 시작하면 시퀀스 데이터에서 질의 시퀀스와 일치하거나 유사한 시퀀스를 찾아 빈도수로 정렬하여 가장 상위에 해당하는 시퀀스를 기준으로 목적지를 추론한다.



[그림 4] 목적지 추론

NHN에서 제공하는 미투데이의 오픈 API인 me2API에 탑재하여 자동 블로깅을 위한 어플리케이션을 구현하였다. 그림 5와 같이 올리기 버튼을 누르게 될 경우 사용자의 입력 없이 자동으로 문장이 생성되어 현재 사용자의 상황을 블로깅하게 된다.

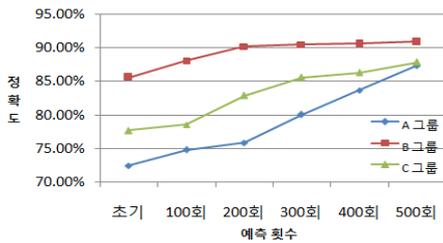


[그림 5] 자동 블로깅 어플 화면

4.2 평가

정확도를 평가하기 위해 3개월간 수집된 6명의 사용자 데이터를 기반으로 500회의 문장을 마이크로 블로그에 블로깅을 하였다. 자동으로 생성된 문장 중 목적지를 추론한 문장은 GPS의 오작동 혹은 사용자가 처음으로 이동하는 장소 등의 요소가 있음에도 시퀀스 매칭의 학습 효과로 최종 평균 90% 이상의 정확도를 보였다. 사용자의 행동이 추론된 문장에서는 휴식과 음악 감상과 같은 유사한 경험으로 인해 목적지보다는 낮은 정확도로 평가되었다. 모든 문장에 대해서 위치 혹은 행동이 경험과 전혀 다르게 추론된 문장은 0점, 행동이 경험과 유사한 상황에 대해서는 사용자가 직접 1점에서 5점 사이의 점수를 입력하도록 하여 평균 88.73%의 정확도를 평가되었다.

그림 6은 비슷한 형태를 보이는 사용자를 3그룹으로 나누어 평가한 것으로, 활동적 성향인 A 그룹은 이동하는 장소가 다양하고 사용하는 어플리케이션이 많아 초기에는 좁은 폭으로 정확도가 상승하였으나, 500회 이상의 학습 데이터를 수집 한 후에는 정확도가 높은 폭으로 상승하였다. 비활동적 성향인 B 그룹은 높은 초기 정확도로 시작하여 100회에서 200회 사이에 상승한 후 평균화되는 것을 볼 수 있다. 같은 비활동적 성향이지만 사용하는 어플리케이션이 다양한 C 그룹은 B 보다는 낮지만, 한번 상승한 후 평균화되는 모습을 보였다.



[그림 6] 정확도 평가

5. 결 론

본 논문에서는 스마트 폰 상에서 사용자의 상황을 자동으로 블로깅 해주는 시스템을 설계하고 구현하였다. 제안한 시스템은 나이브 베이즈 분류기를 이용하여 사용자의 행동을 분류하고 시퀀스 매칭을 이용하여 사용자의 상황을 추론하였다. 그리고 수집된 컨텍스트는 4W를 기반으로 사용자의 상황에 적합한 문장을 자동으로 생성하여 블로깅하도록 하였다.

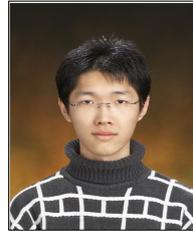
컨텍스트 기반 모델의 성능을 평가한 결과 88.73%의 정확도를 보여 자동 블로깅 서비스가 가능함을 보여주었다. 실험이 계속될수록 추론 기능의 학습효과로 인해 정확도가 높아져 초기 상태보다 우수한 결과를 얻을 수 있음을 확인하였다.

참 고 문 헌

[1] 유혜림 외(2010), 웹 서비스 형태 변화에 따른 소셜 네트워크 서비스의 진화, 한국인터넷정보학회논문지, 11(3), 52-62.

[2] 민준기 외(2010), 모바일 소셜 네트워크 서비스를 위한 사용자의 컨텍스트 확률추론, 정보과학회논문지, 16(3), 361-365.
 [3] 송재구 외(2006), 유비쿼터스 컴퓨팅 정보관리를 위한 컨텍스트 기반의 데이터 모델링, 한국콘텐츠학회논문지, 6(3), 55-62.
 [4] P. Korpipaa, et al.(2003), Bayesian approach to sensor-based context awareness, Personal and Ubiquitous Computing, 7, 113-124.
 [5] 길명선 외(2010). 시계열 데이터베이스에서 선형 추세 제거 서브시퀀스 매칭, 정보과학회논문지, 16(5), 586-590.
 [6] 김용훈(2006), 상황인지 환경에서의 사용자 위치예측 알고리즘에 관한 연구, 석사학위 논문, 연세대학교.
 [7] Statistics Canada, General social survey on time user, <http://www.statcan.na>

박 재 민



- 2010.02 : 공주대학교 컴퓨터공학부 졸업
- 2012.02 : 공주대학교 컴퓨터공학과 석사 졸업
- 관심분야 : 인공지능, 게임 AI
- E-Mail: soulsieg@kongju.ac.kr

이 상 용



- 1984년 2월 : 중앙대학교 전자계산학과 졸업
- 1988년 2월 : 일본동경대학교대학원 총합이공학연구과 공학석사
- 1988년 ~1989년 : 일본 NEC 중앙연구소 연구원
- 1993년 2월 : 중앙대학교 일반대학원 전자계산학과 공학박사
- 1996년 ~1997년 : University of Central Florida 방문교수
- 1993년~현재 : 공주대학교 컴퓨터공학부 교수
- 관심분야 : 인공지능, 컨텍스트 예측, 컴퓨터게임 등
- E-Mail:sylee@kongju.ac.kr