

# 모바일 컨텍스트 로그를 사용한 계층적 이야기 구성 모델

이병길<sup>0</sup> 조성배  
 연세대학교 컴퓨터과학과  
 {byulyi<sup>0</sup>, sbcho}@sclab.yonsei.ac.kr

## A Hierarchical Storytelling Model Using Mobile Context Log

Byung-Gil Lee<sup>0</sup> Sung-Bae Cho  
 Dept. of Computer Science, Yonsei University

### 요 약

휴대폰의 사용영역이 넓어지면서 휴대폰에 저장되는 컨텍스트 정보 활용에 관심이 높아지고 있다. 하지만 정보의 양이 방대하기 때문에 개인이 정보를 분석하여 자신에게 필요한 정보로 바꾸기 위해서는 많은 노력이 필요하다. 본 논문에서는 휴대폰으로부터 컨텍스트 정보를 수집하여 활용할 수 있는 방법으로 개인이 하루 동안 경험한 일에 대한 정보를 한 눈에 알아볼 수 있도록 도와주는 계층적 이야기 구성 모델을 제안한다. 계층적 이야기 구성 모델은 3단계로 구성된다. 우선 각각의 로그를 분석하여 관련 있는 것들을 그룹으로 분류하고 분류된 그룹 내에서 설정된 경로에 대한 가중치를 계산하여 해당 그룹의 가중치로 저장한다. 마지막으로 그룹간의 경로에 대한 가중치를 계산하여 가장 높은 가중치를 갖는 경로를 찾아 이야기 구성 모델로 설정한다. 계층적으로 이야기 경로를 선택한 경우와 그룹으로 분류하지 않고 경로를 계산한 경우의 시간 복잡도를 비교·평가하여 성능을 측정하였다. 이야기 구성모델을 계층적으로 분류했을 때의 성능이 분류하지 않은 경우보다 경로를 선정할 때 더 높은 성능을 나타내었다.

### 1. 서론

최근 휴대폰은 일상생활에서 없어서는 안 될 필수품이 되었다. 사람들은 휴대폰으로 다양한 서비스를 요구하게 되었고 이로 인하여 휴대폰의 성능은 하루가 다르게 발전했다. 최근 휴대폰은 TV를 보고, 사진을 찍고, 음악을 듣는 등 다양한 용도로 사용자들에게 서비스를 제공하면서 기존에 통신 수단으로만 사용한 한정된 기능에서 벗어나 다양한 서비스를 제공해주는 장비로 이용되고 있다. 휴대폰의 사용영역이 넓어지면서 휴대폰에 저장되는 개인 정보의 활용에도 사람들의 관심이 높아지고 이를 기본으로 하는 컨텍스트 정보 활용에 대한 관심 역시 증가하고 있다. 컨텍스트(context)란 사전적 의미로는 문맥이나 환경, 정황 등을 의미하는데 유비쿼터스 네트워킹에서 가지는 의미는 사용자가 처한 환경을 컴퓨터가 인식하는 것부터 시작하여 그 환경 아래에서 사용자의 현재 위치, 행동 및 작업 등에 대한 사용자 정보 값과 그 정보들의 변화를 표현한 모든 정보를 통칭하며 이 정보를 얻어내는 과정을 상황인지(context-awareness)라 한다[1].

하지만 정보의 양이 워낙 방대하기 때문에 개인이 이러한 정보를 모두 분석하여 통계를 내고 유용한 정보로 바꾸기 위해서는 많은 노력이 필요하다. 정보의 양은 점차 증가하고 있고 그러한 정보를 유용한 지식으로 바꾸기 위한 기술이 필요하다. 본 논문에서는 휴대폰으로부터 컨텍스트 정보를 수집하여 활용할 수 있는 방법으로 개인이 하루 동안 무슨 일이 발생했는지에 대한 정보를 보다 손쉽게 알아볼 수 있도록 도와주는 계층적 이야기 구성 모델을 제안한다. 계층적 이야기 구성모델은 기본적으로 사용자의 하루 일과는 다양한 사건이 순차적으로 연결되어 있다는 것과 특정한 시간대별로 어떤 사건들이 발생되는지를 분석하여 동일 시간대에 발생 가능한 사건들을 동일한 그룹으로 구성한다. 구성된 이야기 흐름은 발생된 사건이지만 기록되지 않은 사건을 기억할 수 있도록 도움을 준다.

### 2. 관련연구

최근 휴대폰을 사용하여 컨텍스트 정보를 사용한 연구가 많이 진행되고 있다. 마이크로 소프트 연구소의 경우 사용자의 이벤트 캘린더 정보로부터 중요한 정보를 찾아내는 기법을 연구하고 있다. 사용자는 과거에 일어났던 일을 잊는 빈도수가 증가함에 따라 사용자가 정보를 기억하는 패턴을 분석하고 사용자의 일상 생활을 요약하여 핵심 사항만을 볼 수 있는 브라우저 시스템인 MemoryLens를 개발하였다[2]. 이 과정에서 현재까지는 캘린더 정보로부터 에피소드를 획득하는 방법, 이미지 정보로부터 에피소드 이미지를 찾아내는 방법 등이 활용되고 있으며 데스크탑에서 이용되고 있다.

표 1. 관련연구 비교

	MS Research	Comic Diary
목적	에피소드 탐지	컨퍼런스 경험 요약
데이터	이벤트 캘린더	PDA 정보&센서정보
사용자입력	없음	있음
모델	Bayesian Network	Story Stream
서비스 제공	Desktop PC	Desktop PC
제공 방법	브라우저 형식	만화일기 형식

일본 ATR 연구소는 Comic Diary라는 사용자가 학회에 참가한 기억을 만화 형식으로 보여주기 위한 시스템을 개발하였다. 즉 사용자가 학회에서 경험한 것을 텍스트형식으로 보여주는 것보다는 만화 형식으로 요약하여 보여줌으로써 광범위한 기억을 떠올리게 한다. 이 연구에서 만화는 사전에 구축해 놓은 시나리오를 바탕으로 사용자가 어떤 경험을 했는지를 객관식 형태의 서술지에 답변함으로써 구성된다. 이러한 시나리

오기반 만화 생성은 사전에 정해 놓은 시나리오 범위 내에서 만화 생성할 수 있다는 단점이 있다. 보다 광범위한 일상의 내용을 만화로 담기 위해서는 좀 더 유연한 방법이 필요하다[3].

3. 계층적 이야기 구성 모델

3.1 시스템 개요

본 논문에서 제안하는 모델의 전체 흐름은 그림 1과 같다. 휴대폰으로부터 수집 가능한 컨택트 로그를 항목별로 수집한다. 수집된 로그는 시스템에 적용하기 위해 전처리 과정을 거치는데 이 과정에서는 분리되어 있는 로그를 '일정' 시간 중심으로 통합하고, 통합된 로그에는 베이지안 네트워크로 구성된 모델을 사용하여 각 통합로그별로 가중치를 설정한다[4]. 계층적 이야기 구성 단계에서는 이야기 구성 모델에서 로그별로 해당하는 노드를 탐색하여 매칭하고 해당 이야기 구조를 선택한다. 이야기 흐름 선택이 완료되면 적용되는 어플리케이션에 맞게 정보를 제공한다.

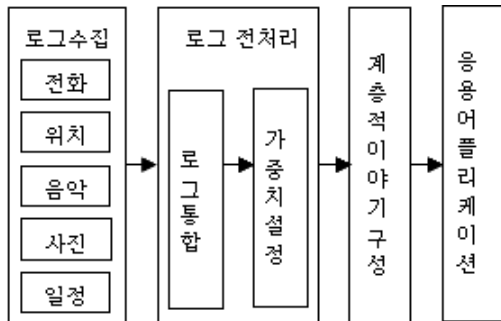


그림 1. 계층적 이야기 구성모델의 전체개요

3.2 로그 전처리

로그 전처리 과정은 로그 통합과 가중치 설정 2단계로 구성된다. 우선 휴대폰을 통하여 수집된 로그를 하나의 로그로 통합한다. 통합된 로그는 시작 및 종료 시각, 장소, 일정, 중요도, 사진, MP3, 전화, 위치정보로 구성된다. 시작시각, 종료시각, 장소, 일정과 중요도는 사용자가 입력한 일정 정보나 위치 정보로부터 채워진다. 이때 일정 정보와 위치 정보의 시간이 중복될 경우 사용자가 입력한 일정 정보가 우선된다.

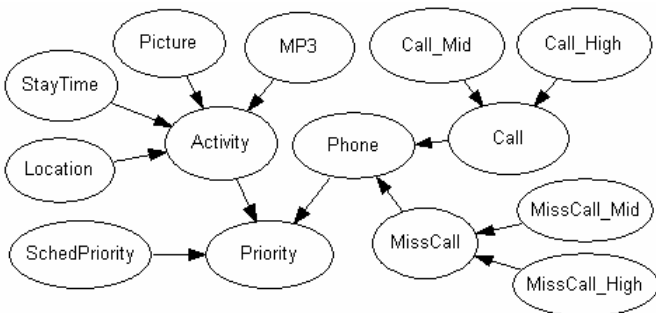


그림 2. 가중치 설정을 위한 베이지안 네트워크 모델

사진, MP3, 전화 정보는 이미 설정된 일정의 시작 시각부터 종료 시각까지 일어난 횟수나 사용된 시간 정보를 통합 로그에 기록한다. 그 다음으로 통합된 로그에 베이지안 네트워크 모델을 통하여 가중치를 부여한다[4]. 본 논문에서 사용한 가중치 모델은 전문가에 의해 설계되어 추론과정을 통해 각 로그에 가중치를 설정한다(그림 2).

3.3 계층적 이야기 구성

이야기 구성 경로 선정은 3단계로 진행된다(그림 3).

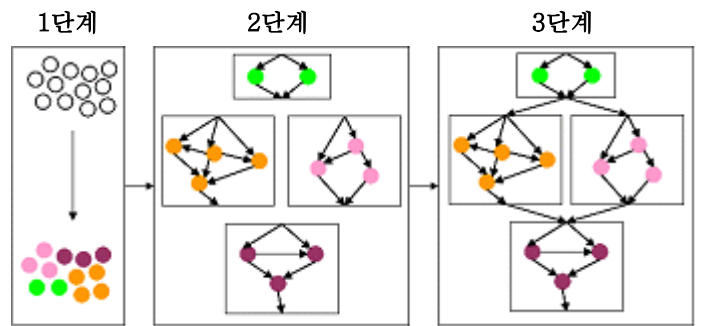


그림 3. 계층적 이야기 구성 모델

첫 단계로 수집된 로그를 각 노드를 생성하여 연관성 있는 노드들끼리 그룹으로 묶어준다. 연관성은 특정한 시간 간격으로 발생할 수 있는 이벤트들에 사진 정의된 라이브러리를 사용하여 파악한다. 각 노드들의 그룹설정이 끝나면 다음 단계에서는 동일한 그룹으로 이루어진 노드들 사이의 경로를 설정하고 설정된 경로 중에서 최대값을 갖는 경로를 탐색한다. 각 그룹 내부의 최대값을 갖는 경로 탐색이 끝나면 그룹 내부 경로의 가중치 합을 그룹의 가중치로 설정하고 내부 경로는 별도로 저장을 해 놓는다. 경로 탐색에 사용되는 방법은 로그 전처리 단계에서 설정한 가중치를 이용하여 그림 4에 정의된 탐욕적 방법(greedy method)을 사용하여 탐색한다[5].

```

procedure StoryStream (N, w, n0)
1. begin
2.    $N_T := \{n0\};$ 
3.   for all  $n \in (N - N_T)$  do
4.     if  $(n0, n)$  exists set  $\{n\} := w(n0, n);$ 
5.     else set  $l[n] := \infty;$ 
6.   while  $N_T \neq N$  do
7.     begin
8.       find a node  $u$  such that
9.          $\{n\} := \max\{l[n] | n \in (N - N_T)\};$ 
10.       $N_T := N_T \cup \{u\};$ 
11.      for all  $n \in (N - N_T)$  do
12.         $\{n\} := \max\{\{n\}, \{u\} + w(u, n)\};$ 
13.      endwhile
14.    end StoryStream
    
```

- \* G: arbitrary connected graph
- \* n0: the initial beginning node
- \* n: the current node
- \* w: the weight of node
- \*  $w(n1, n2)$ : the weight from  $n1$  and  $n2$
- \*  $\{n\}$ : set of node in the graph G
- \*  $N_T$ : the set of all nodes in the graph G

그림 4. 이야기 구성 경로 선정 의사 코드

탐색을 할 때 노드의 수에는 영향을 받지 않고 단지 가중치에만 영향을 받기 때문에 선택된 노드 수가 많아도 가중치가 낮다면 그 경로를 선택하지 않고 가중치가 높은 다른 경로를 선택한다. 마지막 단계에서는 각 그룹 사이의 경로를 설정하고 2단계에서 탐색한 것과 마찬가지로 가중치가 최대값을 갖는 경로를 탐색하여 2단계에서 저장된 경로를 이용, 최종 이야기 흐름 구성을 완성한다. 이렇게 계층적 탐색을 수행하는 이유는 모든 경로를 다 탐색할 경우 탐색시간이 많이 걸려서 효율적인 탐색을 수행할 수 없기 때문이다.

4. 실험 및 결과

4.1 시나리오를 사용한 응용 어플리케이션 적용

계층별 이야기 경로를 선정하기 위해서 주어진 시나리오를 중심으로 컨텍스트 데이터를 수집하여 실험을 진행하였다. 이렇게 수집된 컨텍스트 정보를 이해하기 쉽도록 만화형식으로 표현하였다. 그림 5에서 가로축은 시간을 의미하고 세로축에서 윗부분은 컨텍스트 항목을 의미하는데 각 항목별로 표시가 되어있는 것은 해당 컨텍스트 로그가 수집되었음을 의미한다. 가로축의 아래 부분은 수집된 많은 로그 항목들 중에서 가장치를 계산하고 본 논문에서 제안한 방법을 적용하였을 때 최종적으로 선택된 각 항목들을 하나의 이미지로 변경한 결과이다. 하나의 컨텍스트 로그를 이미지나 다른 어플리케이션에 적용하여 표현을 했을 경우에 사용자들은 높은 관심을 표현했지만, 반대로 제안한 방법을 텍스트 형식으로 보여주는 경우 일부 사용자들은 이야기 구성 모델에 대한 필요성을 직접적으로 느끼지 못하였다.

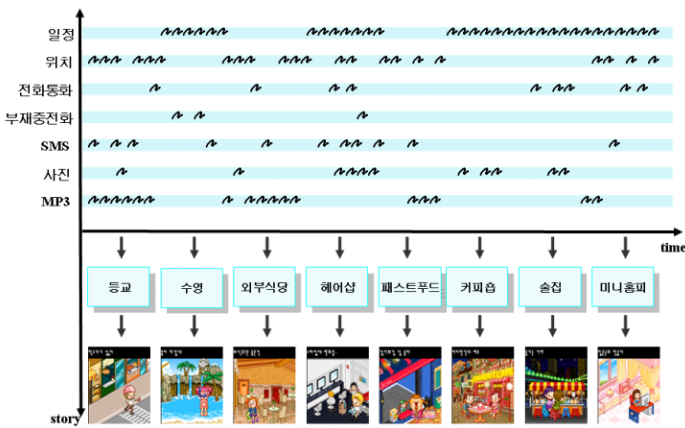


그림 5. 응용 어플리케이션의 예

4.2 시간 복잡도 분석

본 논문에서 사용한 탐욕적 방법은 Dijkstra 알고리즘으로 이야기 경로를 선택할 때 단계를 나누지 않고 모든 노드를 탐색할 할 때 최악의 경우 수행시간은  $O(N^2)$ 이다. 하지만 노드를 계층을 구분하여 단계별로 탐색을 할 경우 수행시간은 그림 6과 같다.

$$\begin{aligned}
 & O(1\text{단계}) + O(2\text{단계}) + O(3\text{단계}) \\
 & = O(1) + [O(n_1^2) + O(n_2^2) + \dots + O(n_m^2)] + O(m^2)
 \end{aligned}$$

- \*  $n_1 + n_2 + \dots + n_m = N$
- \*  $n_1, n_2, \dots, n_m$ : 각 그룹에 포함되어 있는 노드의 수
- \*  $m$ : 그룹 수
- \*  $N$ : 전체 노드의 수
- \*  $O(n_1^2), O(n_2^2), \dots, O(n_m^2)$ : 각 그룹내부 시간 복잡도
- \*  $O(m^2)$ : 그룹에 대한 시간 복잡도

그림 6. 계층별 이야기 경로 선정 시간 복잡도

그러나 최악의 경우  $n=m$ 이라면 계층에 대한 의미는 존재하지 않는다. 이러한 경우를 피하기 위해 사전 정의 라이브러리는 로그간의 연관성과 관련된 많은 양의 데이터를 가지고 있어야 한다.

본 논문에서 시나리오를 통하여 사용한 계층별 이야기 구성 모델에서는 전체 노드수가 38개, 그룹 수가 13개이다. 만약 계층을 나누지 않고 경로를 탐색할 때 시간 복잡도는 최악의

경우  $38^2=1444$ 가 걸린다. 하지만 제안한 방법의 시간 복잡도는 각 그룹 내부의 시간 복잡도와 그룹간 시간 복잡도를 계산해야 한다. 우선 그룹 내부의 시간 복잡도는 각 그룹별로 시간 복잡도를 계산하고 그 값들을 모두 합해주면 된다. 마지막에 추가된 상수 1은 시작그룹은 아무런 노드를가지고 있지 않기 때문에 탐색시간 1이 추가된다. 이와 같은 방식으로 모든 그룹내부의 시간 복잡도를 계산하면 아래와 같다.

$$3^2 + 2^2 + 3^2 + 3^2 + 3^2 + 4^2 + 3^2 + 3^2 + 2^2 + 5^2 + 3^2 + 4^2 + 1 = 129$$

그룹 내부 탐색이 끝나면 그룹간의 시간 복잡도를 계산한다. 그룹수는 시작 그룹(노드)를 포함하여 총 13개로 구성되어 있으므로 그룹간 시간 복잡도는  $13^2=169$ 가 된다. 즉 제안한 방법의 총 시간 복잡도는  $129 + 169 = 298$ 이 소요된다. 따라서 계층을 나누지 않고 탐색하는 경우와 계층별로 탐색하는 경우 최악의 경우만을 비교해 볼 때 계층별로 탐색하는 것이 계층이 없는 경우에 비해서 성능이 약 4.85배정도 좋다는 것을 알 수 있다.

5. 결론 및 향후계획

본 논문에서는 개인의 휴대폰에서 추출된 컨텍스트 정보를 사용하여 개인의 일상 로그를 이야기 구성 모델에 적용해서 하루 일과를 정리할 때 보다 자연스러운 형태로 표현할 수 있는 방법을 제공하였다. 이것은 사용자가 하루 일과를 효과적으로 정리하고 관리할 수 있도록 많은 도움을 준다. 하지만 계층별 이야기 구성모델은 어떤 응용 어플리케이션을 사용하느냐에 따라 표현방식에는 많은 차이를 보이면서 사용자의 만족도에도 차이가 발생하였다.

향후 계획으로는 계층적 이야기 구성 모델을 잘 표현할 수 있는 응용 어플리케이션 개발이 필요하고 다양한 경우에 모두 적용될 수 있도록 계층적 이야기 구성 모델 확장에 대한 연구가 필요할 것으로 보인다.

감사의 글

본 연구는 정보통신부 및 정보통신 연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음, IITA-2005-(C1090-0501-0019).

6. 참고문헌

- [1] A. K. Dey, "Context-aware computing: The cyber desk project," *Proc. of the AAAI 1998 Spring Symposium on Intelligent Environments (AAAI Technical Report SS-98-02)*, pp.51-54, Mar 1998.
- [2] E. Horvitz, S. Duamis, and P. Koch, "Learning predictive models of memory landmarks," *26th Annual Meeting of the Cognitive Science Society*, August 2004.
- [3] Y. Sumi, R. Sakamoto, K. Nakao and Kenji Mase, "ComicDiary: Representing individual experiences in a comics style," *In Proceedings of UbiComp 2002, LNCS 2498*, pp.16-32, Springer, 2002.
- [4] B.-G. Lee, S.-B. Cho, "Learning predictive model of memory landmarks based on bayesian network using mobile context log," *Proc of KCC*, vol. 32, no. 2, pp.550-552, 2005
- [5] M. G. C. Resende and C. C. Ribeiro, "Greedy randomized adaptive search procedures," In F. Glover and G. Kochenberger, editors, *Handbook of Metaheuristics*, pp.219-249, Kluwer Academic Publishers, 2003.