

AP 관계성 분석을 통한 위치 추정

석호식¹⁾⁰, 황규백²⁾, 장병탁¹⁾

¹⁾서울대학교 컴퓨터공학부 바이오지능 연구실

²⁾송실대학교 IT 대학 컴퓨터 학부 기계학습 연구실

hsseok@bi.snu.ac.kr, ²⁾kbhwang@ssu.ac.kr, btzhang@cse.snu.ac.kr

Location Estimation through AP Relationship Analysis

Ho-Sik Seok¹⁾⁰, Kyu-Baek Hwang²⁾, Byoung-Tak Zhang¹⁾

¹⁾Biointelligence Lab., School of Computer Science and Engineering, Seoul National University

²⁾Machine Learning Lab., School of Computing, College of Information Technology, Soongsil University

요 약

본 논문에서는 특정 위치에서 여러 개의 AP(Access Point)에서의 RSS(Radio Signal Strength)를 측정하는 데이터를 획득하였을 때, 과거 측정 데이터와 비교하여 해당 위치를 추정하는 방법을 소개한다. 모든 환경이 동일하다면 과거 데이터의 패턴을 분석하여 현재 위치의 분포를 추정하는 것이 가능하지만, 과거 데이터가 측정된 환경과 현재 데이터가 측정된 환경이 상이하다면, 측정 데이터의 성격이 틀러지기 때문에 위치 추정 작업의 난이도가 높아진다. 본 논문에서는 이와 같이 동일한 위치에 대한 과거 측정 데이터와 현재 측정 데이터의 성격이 같지 않는 상황에서 위치 추정의 정확성을 높일 수 있도록, 과거 데이터를 분석하여 AP 간의 관계성을 분석한 후 AP 관계성을 이용하여 데이터 측정 가능성을 계산하여, 가장 가능성이 높은 위치를 선택할 수 있는 방법을 소개한다. 제안된 방법은 훈련 데이터를 이용하여 각 위치를 설명할 수 있는 AP 관계 모델을 만든 후, Landmark 데이터를 활용하여 만들어진 AP 관계 모델을 변화시키고 이를 활용하여 위치를 예측하는 절차로 구성된다. 제안 방법론은 2007 IEEE ICDM(International Conference on Data Mining) Data Mining Contest의 데이터에 적용되었으며, 과거 컨테스트 참여자들의 실험 방법론과 비교해 보았을 때 17위에 해당하는 성능을 보였음을 확인하였다.

1. 서 론

모바일 테크놀로지의 발전과 함께 AP(Access Point)가 건물 내부에 설치되면서 사용자가 소지한 모바일 기기에서 Wi-Fi 신호를 수신하여 건물 내부에서 위치를 확인할 수 있는 가능성이 소개되었다. 위치 추정을 위해서 흔히 사용되는 방법은 RSS(Radio Signal Strength)를 이용하는 것으로 수신한 RSS를 이용하여 위치를 추정한다. 그러나 문제는 Wi-Fi 신호가 온도, 습도, 측정 시기와 같은 환경의 영향에 매우 민감하다는 것이다. 따라서 과거에 측정된 RSS 데이터에 기반하여 현재 RSS 데이터의 측정 위치를 추정하는 것은 매우 어려운 작업이 된다[1]. 이와 같은 문제를 해결하기 위해 사용할 수 있는 기법이 기계 학습(Machine Learning) 기법을 사용하는 것으로, 과거 데이터의 분포를 분석하여 AP간의 관계성을 분석·학습한 후 이를 이용하여 위치를 추정하는 방법을 고려할 수 있다.

이와 같은 추정 방식은 통계적 관계성을 분석하여 네트워크 구조를 획득한 후 이를 이용하여 추정 작업을 수행하는 방법론[2]과 유사한 것으로 본 논문에서는 AP간 관계성을 분석한 후, 획득한 AP 관계성 정보를 위치 추정에 사용한다. 이와 같이 먼저 AP 관계성 정보를 분석한 후 이 관계성을 이용하여 위치를 추정하는 것은 그림 1에서 볼 수 있는 특성 때문이다. 그림 1에서 보이고 있는 것은 2007 IEEE ICDM(International Conference on Data Mining) Data Mining Contest 데이터가 측정된 환

경으로, AP 데이터 측정에 사용된 위치가 표시된 지도의 일부이다.

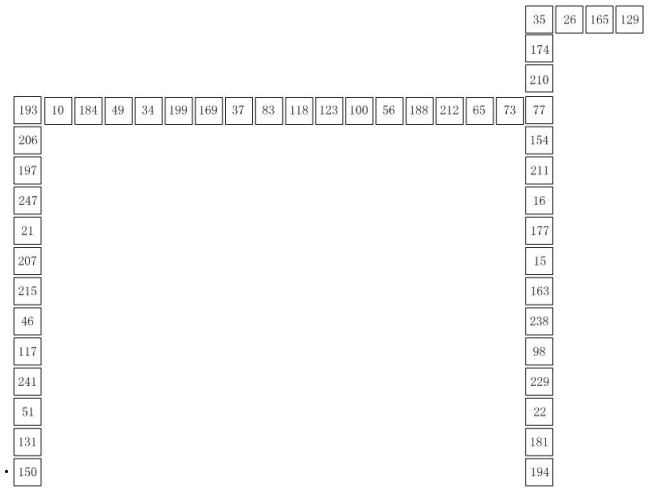


그림 1 데이터가 측정된 Map의 일부

그림 1에서 알 수 있듯이 측정된 위치의 인덱스 혹은 위치 레이블(위치1, 위치 2와 같은 위치 레이블)이 고정되어 있다면 특정 위치(예를 들어 위치 118)에서 측정할 수 있는 AP가 고정된다. 따라서 위치 레이블이 함께 제공되는 AP 데이터를 분석할 경우, 특정 위치 주변에서 활성화되어 있는 AP 인덱스를 확보하게 되며, 이 정보를

이용하면 위치 레이블을 알려져 있지 않은 데이터의 위치를 추정하는 것이 가능하다.

그러나, 문제가 되는 것은 Wi-Fi 신호가 환경의 변화에 매우 민감하다는 것이다. 2007 IEEE ICDM Contest에서는 문제의 난이도를 높이기 위하여 훈련 데이터가 측정된 시기와 테스트 데이터가 측정된 시기를 다르게 하여 전혀 다른 AP 데이터 분포를 이용하여 위치 추정 문제를 소개하였다[3]. 이와 같이 훈련 데이터의 성격과 테스트 데이터의 성격이 상이한 상황은 Transfer Learning에 대한 연구에서 활발하게 다루어지고 있는 상황이다[4, 5]. Transfer Learning에 대한 관련 연구와 데이터 측정 환경의 특성을 동시에 고려하였을 때 얻을 수 있는 결론은 데이터 위치 추정을 위해서는 환경 변화에 관련 없는 변하지 않는 특성을 찾아야 한다는 사실이다.

본 논문에서는 2007 IEEE ICDM Contest에서 소개한 위치 추정 문제를 해결하기 위하여 AP 데이터의 관계성을 분석하여 위치 레이블과 AP 데이터, AP 데이터 간 관계성을 찾고 이를 이용하여 위치를 추정하는 방법론을 소개한다. 본 논문에서 소개하는 방법론은 2007 IEEE ICDM Contest 문제에 적용되었으며, 컨테스트 참여자의 결과와 비교하였을 때 17위에 해당하는 성능을 보임을 확인하였다.

2. 2007 IEEE ICDM Data Mining Contest

2007 IEEE ICDM Data Mining Contest[3]은 두 개의 태스크로 구성된다.

표 1 2007 IEEE ICDM Data Mining Contest Task

Task 1	<i>Indoor Location Estimation</i> Wi-Fi AP에서 수신한 RSS(Radio Signal Strength)를 이용하여 각 측정 데이터의 위치를 추정
Task 2	<i>Transferring the Learned Knowledge for Indoor Location Estimation</i> Task 1과 유사하나 훈련 데이터가 측정된 시기와 테스트 데이터가 측정된 시기가 틀림

표 2 Task 2 데이터 속성

1. AP의 수: 99
2. Training Data: 2322 collections of RSS values
3. Test Data: 3128 collections of RSS values
4. Landmark Data: 53 collections of RSS values
5. Location의 수: 247

표 1과 2에서 컨테스트 데이터에 대하여 설명하고 있다. IDCM Contest는 두 개의 태스크로 구성되어 있다. 태스크 1은 동일 환경에서 수신한 훈련 데이터와 테스트 데이터를 이용하여 테스트 데이터가 측정된 위치를 추정하는 것이다. 태스크 2는 Transfer Learning의 개념을 도입한 것으로 훈련 데이터와 테스트 데이터가 측정된 환경에 차이가 있다. 단 테스트 데이터의 추정 작업에 도움을 제공하기 위하여 테스트 데이터와 동일한 환경에서 측정된 Landmark 데이터를 제공하여 분석 과정에 사용할 수 있도록 한다. 표 2에서 알 수 있듯, 추정해야 하는 위치의 수 247에 비하여 제공되는 Landmark data의 수가 53개에 불과하기 때문에 2322개의 Training data instance를 최대한 활용해야 테스트 데이터의 추정이 가능하다.

3. 데이터 분석을 통한 AP간 관계성, 위치 레이블—AP 관계성 분석

훈련 데이터와 테스트 데이터의 측정 환경이 차이가 있지만, 그림 1에서 알 수 있듯 몇 가지 사항은 변하지 않는다. 첫째, 데이터를 측정하는 위치는 변하지 않으며 둘째, AP의 위치 역시 변하지 않는다. 따라서 훈련 데이터의 분석을 통해 AP의 관계성 및 위치 레이블—AP 관계성을 찾아낸다면, AP 위치에 대한 데이터를 추정할 수 있게 되고 이를 통해 테스트 데이터의 추정이 가능하다. 이는 [2, 6]등의 연구에서 활발히 연구된 것으로 그래프로 표현된 데이터 문제 공간에서 링크를 분석하여 데이터 인스턴스 간 혹은 데이터를 구성하는 특성 간 관계성을 찾고 이를 통해 다른 태스크를 수행하고자 하는 접근법과 궤를 같이 하는 것이다. 이를 보다 자세히 표현하면 다음과 같다.

주어진 테스트 데이터를 Test 데이터, 훈련 데이터 분석을 통해 확보한 각 위치-레이블 설명 모델을 Model이라고 했을 때, 태스크 2에서 요구하는 작업은

$$P(m|Test) = P(Test|m)P(m)/P(Test)...(1)$$

식 (1)을 최대화하는 Model 즉

$\text{argmax}_{m \in \{MODEL\}} P(m|Test)$ 을 만족하는 m 을 찾는 것이다(여기서 MODEL은 247개의 위치 데이터에 관하여 생성된 설명 모델을 의미). 식 (1)을 계산하는 과정에서 $P(Test)$, $P(m)$ 의 성격에 대하여 특별한 가정을 하지 않는다면 식 (1)의 계산에서 가장 중요한 것은 Likelihood Probability $P(Test|m)...(2)$ 가 된다. 식 (2)의 계산을 위해 필요한 것은 환경 변화에 따라 측정 데이터가 어떻게 변화하는지 파악하는 것이다. 그림 2에서 환경 변화에 따른 AP 측정 결과의 변화를 보이고 있다. 그

림 2에서 원은 측정된 AP를 의미하며 링크는 AP간 관계성을 의미한다. 그림 2에서와 같이 측정 환경이 변화하면 다음의 변화를 관찰할 수 있다.

- a) 기존 AP 사라짐: 그림 2 상단의 11번 AP가 그림 2 하단에서는 관찰되지 않는 것처럼 측정 환경 변화에 따라 관찰되었던 AP 라디오 신호가 사라지는 현상이 발생한다.
 - b) 새로운 AP 등장: 그림 2 상단에는 없었던 13번 AP가 그림 2 하단에서 관찰되는 것처럼 환경 변화에 따라 새로운 AP가 관찰되는 현상이 발생한다.
 - c) 측정값의 변화: 그림 2 상단과 그림 2의 하단을 비교하면 3번 AP와 5번 AP를 연결하는 링크의 굵기가 변화한다. 이와 같이 측정 환경의 변화에 따라 측정된 AP 라디오 신호의 세기가 변화한다.
- a)~c)의 현상을 최대한 활용해야만 식 (2)의 효율적인 계산이 가능하다.

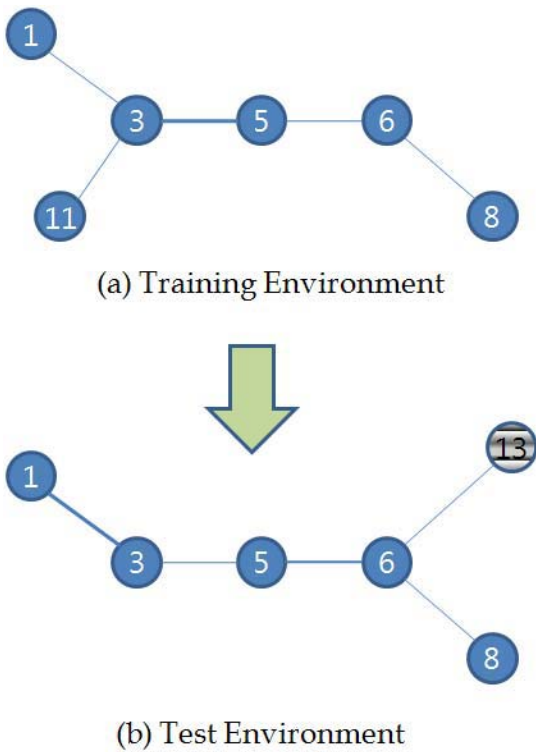


그림 2 환경 변화에 따른 측정 결과 변화

그림 2에서 관찰된 a)~c) 현상을 반영하여 식 (2)를 다시 표현하면 다음과 같다.

$$P(\text{Test}|m) = \underset{m \in \{\text{MODEL}\}}{\text{argmax}} \prod_t \Psi(i, j|m) \Lambda(\text{Te}_k|m) \dots(3)$$

식 (3)에서 t 는 Model을 표시하는 인덱스이며, $\Psi(i, j|m)$ 은 특정 Model에서 AP_i 와 AP_j 의 관계성을 표시한다.

$\Lambda(\text{Te}_k|m)$ 은 모델 m 에서 k 번째 AP가 등장할 확률 혹은 사라질 확률을 나타낸다. 따라서 식 (3)이 의미하는 것은 다음과 같다. AP가 유지될 확률, 등장할 확률, 사라질 확률을 계산하며, 각각의 경우에 대하여 각 AP에 대한 라디오 신호의 세기를 반영하여 주어진 테스트 데이터를 가장 잘 설명할 수 있는 위치 모델을 선택한 후 선택된 위치 모델이 해당 테스트 데이터의 위치라고 예측한다.

표 3 위치 예측을 위한 제안 알고리즘

Step 1: Data Analysis

Input:: Training Data, Landmark Data

Output: Model m

- Link Analysis

Obtain link information from training data.

(Link information represents AP proximity computed using AP RSS)

Modify link distribution using Landmark data.

Compute model m for all locations.

Step 2: Location Estimation

Input:: Test Data instance, Model m

Output: Estimated location for the given test data instance

Compute equation (3) using m

표 3에서 위치 예측을 위한 제안 알고리즘을 설명하고 있다. 식 (3)의 계산을 위하여 제안 알고리즘은 데이터 분석 단계 및 위치 예측 단계로 계산된다.

데이터 분석 단계는 새로운 AP가 등장할 확률, 기존의 AP가 사라질 확률을 계산하고 환경 변화에 따른 AP 관측값 변화(RSS 변화)를 구하기 위한 것으로 훈련 데이터를 분석하여 링크간 관계성을 계산하는 것이 주목적이다. 데이터 관계성은 RSS를 이용하여 계산되며 가능한 AP pair를 생성한 후 생성된 AP pair의 RSS 관계를 구하여 해당 AP pair의 유의미성을 수치로 표현한다. AP가 등장하거나 사라질 확률은 Landmark 데이터를 이용하여 계산되며, Landmark 데이터의 정보를 이용하여 훈련 데이터에서 계산된 AP pair 분포를 변화시킨다. 이때 AP pair를 구성하는 관찰 데이터는 정규 분포를 따른다고 가정하며 Landmark 데이터의 관측값을 이용하여 정규 분포의 인자를 변형하도록 한다.

위치 예측 단계는 데이터 분석 단계에서 구해진 위치 모델을 이용하여 주어진 테스트 데이터 인스턴스의 위치를 추정하는 단계로 식 (3)을 계산하는 과정으로 이루어진다.

4. 예측 결과

식 (3) 중 $\Lambda(Te_k|m)$ 만을 반영하여 ICDM Data Mining Contest 데이터의 태스크 2에 대하여 예측 작업을 진행하여 보았다. 예측 결과는 다음과 같다.

표 4 예측 결과 비교

Rank	Precision($\frac{\text{Number of correct estimation}}{\text{Number of test data}}$)
1	0.322
2	0.314
3	0.313
4	0.311
5	0.297
6	0.296
7	0.294
8	0.2928
9	0.2921
10	0.289
11	0.279
12	0.252
13	0.25
14	0.243
15	0.174
16	0.173
17	0.162(제안 알고리즘)
18	0.153

식 (3) 중 $\Lambda(Te_k|m)$ 만을 반영하여 위치 예측을 한 결과 2007 Data Mining Contest 참여자 중 17위에 해당하는 예측 성능을 보였다.

5. 결론 및 추후 연구

본 논문에서는 Transfer Learning의 아이디어를 활용하여 2007 Data Mining Contest의 태스크 2를 해결할 수 있는 알고리즘을 제안하였다. 제안 알고리즘은 주어진 훈련 데이터를 이용하여 AP관계 모델을 생성한 후, Landmark 데이터를 활용하여 변화된 환경에 맞게 AP 관계를 변화시키고 변화된 모델을 위치 예측에 사용하는 절차를 따른다. 고려해야 할 두 가지 관계 Ψ, Λ 중 Λ 관계 한 가지만을 구현하여 성능을 확인한 결과 경쟁력 있는 예측 성능을 보임을 확인할 수 있었다.

본 연구의 발전 과정에서는 Ψ 관계를 반영한 모델을 생성하여 예측 성능을 향상시키는 연구를 진행하고자 한다.

참고 문헌

[1] Yang, Q., Pan, S. J., and Zheng, V. W., Estimating Location Using Wi-Fi, *IEEE Intelligence Systems*, 23(1), 8-13, 2008.
 [2] Rattigan, M. J., Maier, M., and Jensen, D., Exploiting Network Structure for Active Inference

in Collective Classification, Proceedings of the Seventh IEEE International Conference on Data Mining Workshops, 429-434, 2007.
 [3] <http://www.cse.ust.hk/~qyang/ICDMDMC07>
 [4] Caruana, R., Multitask Learning, *Machine Learning*, 28, 41-75, 1997.
 [5] Marx, Z., Rosenstein, M. T., and Kaelbling, L. P., Transfer Learning with an Ensemble of Background Tasks, NIPS Workshop on Inductive Transfer, 2005.
 [6] Popescul, A and Ungar, L. H., Statistical Relational Learning for Link Prediction, Proceedings of the Workshop on Learning Statistical Models from Relational Data at IJCAI-2003, 2003.