

인간 이동 데이터와 BFI 성격 데이터를 이용한 인간의 위치 예측

김승연*, 이은별, 송하윤
홍익대학교 컴퓨터공학과

e-mail : brdosa@naver.com, qyftmxk@hanmail.com, hayoon@hongik.ac.kr

Next Location Prediction Through Positioning Data and Big Five Inventory

SeungYeon Kim* Eun Byul Lee, Ha Yoon Song
Department of Computer Engineering, Hongik University

요 약

인간은 성격에 따라 이동패턴이 변화한다고 한다. 이런 점에서 인간의 성격 데이터를 이용하면, 인간의 행동 패턴을 유추해 낼 수 있다. 우리는 실제 실험자들의 GPS데이터와 BFI성격 데이터를 수집하고, Back Propagation Network를 이용하여, 새로운 위치 데이터를 추론하는 과정을 설명하였다. 논문의 내용은 다음과 같다. 첫 번째로 BFI(Big-Five Inventory) 성격평가에 대해 설명한다. 두 번째로 GPS데이터와 성격 데이터를 실험에 적절한 형태로 변환하는 방법에 대해 언급하고, 세 번째로 변환된 데이터를 이용하여 사람의 새로운 위치 정보를 추론할 것이다. 마지막으로 해당 실험의 결과 및 분석 그리고 앞으로의 연구 방향에 대해 언급할 것이다.

1. 서론

이 논문에서 우리는 실제 실험자의 GPS 위치 데이터와 성격 데이터를 이용한 사람의 성격 데이터를 사용할 것이다. 그리고 Back Propagation Network 알고리즘을 이용하여 새로운 위치 데이터를 추론하는 방법론을 보일 것이다 [1].

선행 심리학 논문에 의하면 인간의 이동패턴은 각 개인의 성격에 영향을 받는다고 한다. 예를 들면 외향적인 인간은 내향적인 인간보다 외출을 선호할 것이다[2].

인간의 성격을 나타내는 심리학적 표현 방법은 많이 있지만, 우리는 성격 표현방법에 관하여 가장 일반적으로 많이 쓰이는 BFI(Big Five Inventory)[3]를 이용할 것이다.

이 연구가 다른 위치 예측 방법론과의 차이점은, 첫째로 우리는 예측 알고리즘을 다른 방법론들은 주로 Markov Model을 사용한다 반해[4], 이 논문에서는 Back Propagation Network를 사용한다 있다. 이 의미는 시간에 따른 위치 이동을 Sequence로 본 게 아니라 연속적인 흐름으로 본다는 것을 나타낸다. 두 번째로 우리는 예측하는 방법에 대해 성격 데이터를 사용하는데, 다른 예측 방법론들은 성격 데이터를 사용하지 않는다. 이는 이 논문은 '위치 데이터의 이동에는 사람의 성격이 영향을 준다.'라는 가정을 했기 때문이다[5].

다음 2장에서는 앞서도 언급했던 성격요소인 BFI에 대한 간단한 설명을 할 것이며, 3장에서는 BFI를 이용한 행동 예측 알고리즘의 구성 설명과, 변수의 결정 과정, 그 실험 및 결과를 보여줄 것이다. 그리고 마지막 장에서 해당 연구의 개선 방향과 연구의 방향에 대해서 언급하게 될 것이다.

2. BFI(big five inventory)

BFI데이터는 사람의 성격데이터를 수치화해서 나타낸다. 때문에 이 특징은 컴퓨터 알고리즘에 쉽게 데이터를 적용할 수 있도록 돕는다.

성격의 개인차를 나타내기 위해 성격심리학자들은 다양한 이론을 고안, 연구해왔다. 그 중 McCrae 와 Costa가 주장한 이론을 BFI라 한다.

BFI는 결과를 개방성(Openness), 성실성(Conscientiousness), 열정적(Extraversion), 동조성(Agreeableness), 신경성(Neuroticism) 같은 5가지 요소로 평가하고 있다.

○ Openness to experience (inventive/curious vs. consistent/ cautious) : 경험에 대한 개방성. 예술에 대한 창의성과 지적인 호기심을 가리킨다. 모험, 예술, 호기심 등이 영향을 받는다, 경험에 대한 개방성이 아니라 '지식'으로 보기도 한다.

○ Conscientiousness (efficient/organized vs. easygoing/ careless): 성실성. 세운 계획을 달성하기 위한 노력, 꾸준한 노력하는 인간의 성향을 가리킨다.

○ Extraversion (outgoing/energetic vs. solitary/reserved): 열정적. 에너지가 넘치는 인간, 말이 많은 인간, 긍정적인 사고를 가진 인간의 성향을 가리킨다.

○ Agreeableness (friendly/compassionate vs. cold/unkind): 동조성. 남들과 함께 일을 하는 인간의 성향을 가리킨다.

○ Neuroticism (sensitive/nervous vs. secure/confident): 신경성. 자극에 대해 민감한 정도를 나타낸다. 마음의 상처를 잘 받거나, 외출을 꺼려하는 인간의 성향을 가리킨다.

이런 5가지 요소를 이용하여 인간의 성격을 수치화 시켜 나타낸다. 이 값은 해당 성격 점수가 낮을수록 0에 가깝고 높을수록 5에 가까운 점수를 준다.

3. 실험 설계

논문에서는 두 가지 실험을 할 것이다. 실험자의 위치 데이터를 하나의 BPN에 학습시키고, 학습된 BPN에 각 실험자의 결과와 입력했던 위치데이터간의 개형을 비교하고, 신뢰성에 대해 언급할 것이다.

실험에 사용할 데이터는 대학생 5명분의 6개월간 수집한 데이터를 사용할 것이다. 데이터는 GPS데이터로 수집되었고, 각 위치에 대하여 클러스터로 분석하여 사용하였다. <그림1>은 실제로 실험에 사용할 데이터를 KML 형식을 사용하여 나타낸 것이다. 자세히 설명 하면, 각 클러스터는 실험 대상자의 수집된 '특정 위치'를 나타내며, 각 클러스터가 가지는 원의 크기가 클수록 주로 가는 위치를 나타낸다. 이 원의 크기가 클수록, 즉 자주 가는 곳일수록, 작은 숫자를 부여했다. 예를 들어 클러스터 0은 가장 큰 원을 가지게 된다. 따라서 클러스터 0은 다른 클러스터에 비해 가장 자주 가는 곳을 나타내게 된다.

선행 연구에 따라 '사람이 자주 가는 곳은 그 사람의 성격에 영향을 받는다.'[2]라는 가정을 사용한다. 따라서 자주 가는 곳을 알기 위해, 해당 클러스터의 의미를 파악할 필요가 있다. 위 그림[1]에서는 클러스터 0은 실험자의 '집'을 상징하고, 클러스터 1은 '학교'를 상징한다. 우리는 우선 클러스터 0과 1에만 의미를 부여할 것이다. 그 이유는 첫째, 방법론을 말하는 것이므로, 실험을 복잡하게 만들 필요가 없기 때문이고, 두 번째로 클러스터 0과 1은 전체 이동 클러스터 분포의 80% ~ 90%를 차지한다. 즉 사람의 이동하는 장소의 대다수는 1~2개의 장소에 나타난다 [4]. 따라서 나머지의 클러스터는 '희귀' 확률로 간주해서 클러스터 0,1 이외에는 의미를 부여할 필요가 없다.

<표 1>은 이 실험자의 성격 데이터를 나타낸다. 실험자의 성격의 점수가 클수록 해당 성격의 성향이 강하다는 것을

나타내는데, 우리가 예시로 사용한 실험자의 성격은 첫 번째 줄의 데이터가 된다. 이 실험자의 경우 협조성에 4.22, 신경성에 1.7을 보여주고 있다. 따라서 실험자는 평균의 사람보다 협조성이 좋으며, 덜 민감한 사람이라는 것을 알 수 있다.

<표 2>는 실험자중 한명의 시간에 따른 위치 데이터 값을 나타낸 그래프이다. 이 그래프를 해석하면, 실험자는 새벽에 학교와 집에 있을 확률이 비슷하며, 오후에 집에 있을 확률이 아주 높음을 보여준다. 특이한 점은 이 실험자의 위치 데이터 값이 17시부터 23시 까지 '집'에 있을 확률이 높다. 하지만 20시의 경우 집에 있을 확률이 순간적으로 줄어든다. 이를 미루어 보아 실험자는 20시에 잠시 집을 벗어나 집이나, 학교 이외의 장소에 간다고 생각할 수 있다. <그림 2>는 <표 2>를 그래프 모양으로 나타낸 것이다.

4. 실험

실험의 과정은 <그림 2>와 같은 시간위치 데이터를 하나의 BPN에 학습시키고, 학습된 BPN에 실험자의 정보와 알고자하는 시간을 입력하면 입력한 시간에 대한 위치 데이터 확률 값을 반환하게 한다. 이렇게 반환된 확률 값을 학습하기 전의 위치데이터 값과 비교하여, 실험의 신뢰성을 언급할 것이다. 여기서 실험자의 정보는 성격 데이터가 된다. 즉 성격 데이터의 값이 해당 실험자를 나타낸다고 보면 된다. 성격 데이터를 해당 실험자를 나타내는 값으로 사용할 수 있는 이유는, BFF의 각 요소의 값은 서로 직교성을 만족하기 때문이다[6][7].

실험을 하는 과정에서 모든 실험자의 정보를 하나의 BPN에 입력해야 의미가 있다. 그 이유는 BFF점수를 추론하는 과정의 변수로서 이용하기 위함이다. BPN의 학습



(그림 1) 실제 실험자의 위치정보를 KML 형식으로 나타낸 그림

Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
3.75	3.25	3.2	4.2	1.7
3.6	3.3	2.75	3.2	2.75
3.7	3.3	3.75	2.6	4
4.2	4.33	3.5	3.5	2.6
3.3	3.8	3.25	3.6	2.6

<표 1> 실험자의 실제 BFF 성격 데이터.

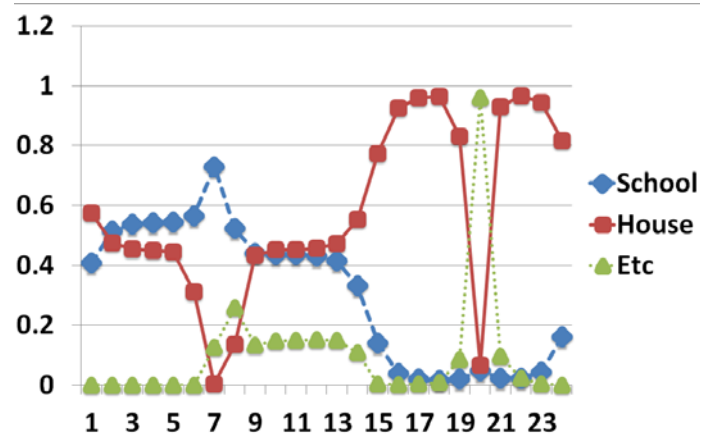
과정은 일종의 입력 값에 대한 결과 값을 나오게 해주는 함수를 만드는 과정이라 볼 수 있다. 만약 한 사람의 정보만을 BPN에 학습 시키면 입력된 성격 데이터의 종류는 하나에 불과 하기 때문에 성격의 영향을 받지 않는 BPN이 나온다. 하지만 만약 복수의 데이터를 입력하면, 여러 성격에 따른 위치 데이터를 학습하게 되므로, 성격의 영향을 받는 BPN을 만들 수 있다.

<그림 3>은 <그림 2>와 같은 시간위치 데이터를 BPN에 학습한 결과이다. 또한 <표 3>는 <그림 3>의 수치 값을 나타낸 것이다. <그림 3>에 대하여 설명하자면, 21시의 경우 해당 실험자의 집에 대한 위치 데이터 값은 0.92이다. 따라서 해당 성격을 가진 사람은 21시에 집에 있을 가능성이 크다고 생각할 수 있다. <그림 2>와 비교하면,

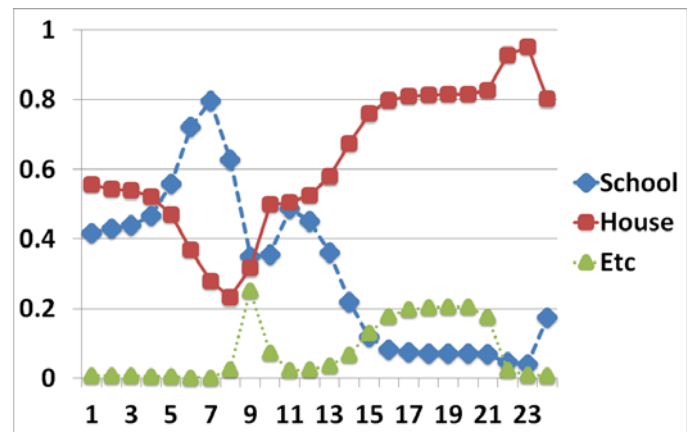
<표 2> 실험자의 시간에 따른 GPS값.

시간	학교	집	기타
0	0.4079	0.575	1.60E-06
1	0.5154	0.4748	4.70E-07
2	0.5378	0.4542	3.65E-07
3	0.542	0.4501	3.49E-07
4	0.5433	0.4461	3.70E-07
5	0.5640	0.3119	1.60E-06
6	0.7297	0.0027	0.12581
7	0.5229	0.1346	0.258757
8	0.440	0.4342	1.33E-01
9	0.4365	0.4512	1.46E-01
10	0.4354	0.4527	1.51E-01
11	0.4319	0.4558	1.51E-01
12	0.4142	0.4721	1.49E-01
13	0.3317	0.5528	1.07E-01
14	0.1404	0.7726	4.00E-03
15	0.0377	0.9262	0.000974
16	0.0188	0.9603	0.003762
17	0.0153	0.9651	0.007591
18	0.0204	0.8301	0.08462
19	0.047	0.0661	0.961264
20	0.0224	0.9302	0.096529
21	0.0234	0.9675	0.021513
22	0.0434	0.9439	0.003479
23	0.1598	0.8169	6.09E-05

<그림 3>의 그래프 개형이 다소 바뀌었는데, 그 이유는 BPN에 학습하는 과정에서 해당 실험자의 위치 데이터 뿐만 아니라, 다른 사람의 위치 데이터도 같이 학습했기 때문이다. 그 결과 그래프의 개형은 BPN이 해당 성격에 대한 패턴을 출력한다. 때문에 실제 데이터하고 비슷하지만, 다른 모양이 나오게 된 것이다. 하지만 그래프의 의미는 크게 다르지 않다. 예를 들어 오후에는 전반적으로 '집'에 있다는 사실과, '집', '학교'이외의 장소는 잘 안 간다는 것을 알 수 있다.



(그림 2) 실제 시간에 따른 위치 데이터 클러스터 확률 분포도



(그림 3) BPN에 학습후의 시간에 따른 위치 클러스터별 확률 분포도.

5. 결론 및 분석

우리는 실험자의 실제 성격 데이터와 위치 데이터를 사용하여, 시간에 따른 사람의 현재 위치를 추론하는 방법론에 대해 언급했다. 또한 우리가 사용한 BPN에 대하여 성격 요소를 변수로 작용하게 하여, 위치 데이터와 성격 데이터간의 관계를 파악하는 방법에 대한 연구를 했다. 이런 실험들을 통하여 실제 위치 데이터를 통한 새로운 위치 데이터를 추론하는 게 가능하다는 것을 보였고, 성격 데이터를 이용하여 특정 위치와 성격간의 관계를 파악할 수 있음을 보였다.

하지만 실험의 결과가 다소 정확도가 떨어짐을 파악할 수 있다. 그 이유는 원천 데이터의 절대수가 부족하기 때문이라 생각한다. 따라서 우리의 다음 실험은 원천 자료를 더욱 수집하여, 실험의 정확도를 유의미한 수준까지 높이는 것이 목적이다.

그 외에도 우리는 심리데이터를 BFF로 사용하였는데, BFF이외의 심리 지표를 심리 데이터로 사용하는 방법에 대해 고려해 볼 것이다. 또한 현재 HMM를 이용한 위치 클러스터 분석법이 있는데, 이런 분석 방법과 논문에서 언급한 예측 방법을 융합하여 한 단계 높은 결과를 내보려고 한다.

6. Acknowledgement

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2012R1A2A2A03046473)

<표 3> 실험자의 시간_위치 데이터를 학습 시킨 위치 데이터 결과 값.

시간	학교	집	기타
0	0.417165	0.556133	0.007005
1	0.43052	0.543914	0.006932
2	0.438452	0.539206	0.006724
3	0.466897	0.52278	0.006007
4	0.557612	0.470128	0.004214
5	0.72152	0.368525	0.002094
6	0.796952	0.278776	0.002024
7	0.627015	0.231363	0.02507
8	0.349829	0.316256	0.250943
9	0.354993	0.498927	0.071594
10	0.486974	0.504547	0.021568
11	0.451416	0.526281	0.024476
12	0.360698	0.580117	0.035008
13	0.21858	0.674623	0.067177
14	0.117705	0.760902	0.129282
15	0.082802	0.799561	0.17863
16	0.073642	0.811099	0.197603
17	0.071335	0.81417	0.202984
18	0.07073	0.815066	0.204424
19	0.0705	0.815721	0.204072
20	0.068174	0.827769	0.176528
21	0.044982	0.929674	0.022286
22	0.039272	0.951559	0.008383
23	0.174208	0.802795	0.007448

참고문헌

[1] Jillian Anable, "'Complacent Car Addicts' or 'Aspring Environmentalists'? Identifying travel behaviour segments using attitude theory", Transport Policy, Vol.12, pp.65-78, 2005

[2] Giuseppe Carrus, Paola Passafaro, Mirilia Bonnes "Emotions, habits and rational choices in ecological behaviours: The case of recycling and use of public transportation" Journal of Environmental Psychology,2008.

[3] Walter Mischel, 손정락 역, 성격심리학, 시그마프레스 출판사, ch-3, pp.273-293, 2006

[4] Burbey, Ingrid E. Predicting future locations and arrival times of individuals. Diss. Virginia Polytechnic Institute and State University, 2011.

[5] Schmitt D. P., Allik J., McCrae R. R., Benet-Martinez V. "The geographic distribution of Big Five Personality Traits: patterns and profiles of human self-description across 56 nations", J. Cross Cult. Psychol., Vol.38, pp.73-212. 2007.

[6] Jack Block, "A Contrarian View of the Five-Factor Approach to Personality Description", University of California, Berkeley, Psychological Bulletin, Vol. 117, No. 2, 187-215, 1995.

[7] Geoffrey Hinton, "NIPS Tutorial on: Deep Belief Nets", Canadian Institute for Advanced Research & Department of Computer Science University of Toronto, 2007.