



An Interpretation and Translation System Using Location Based Big Data Analysis

Sangjin Lee¹, and Seokhoon Kim²

¹Department of Hospitality & Tourism, Changshin University

²Department of Computer Software Engineering, Soonchunhyang University

ABSTRACT

Nowadays, an accuracy of interpretation and translation system, which are based on a natural language, are quotidianly increasing, because a big data processing method has been dramatically progressed. However, there are still some lacks for an user who is needed a colloquial style interpretation and translation. To this end, we propose an interpretation and translation system for the user, which are using a location information based on big data analysis. When the proposed system collect data, the system is preferentially gathering the data which are generated in the near regions, and which are related with the regions. At this time, the criteria of data gathering process are each data location information. And then, the system will configure various data set by their data type. The configured data set, which it will be used pre-learning parameters in a deep neural network. Based on these parameters for the pre-learning process, the proposed system carry out a deep learning process. By doing this, the system provide a improved interpretation and translation result for colloquial style. To verify an excellence of the system, we compare with an existing system. In the performance evaluation, the proposed system show enhanced results about 10% than the existing system.

© 2017 KKITS All rights reserved

KEYWORDS : Interpretation, Translation, Location information, Bigdata, Natural language, Artificial intelligence

ARTICLE INFO: Received 12 January 2017, Revised 4 April 2017, Accepted 7 April 2017.

*Corresponding author is with the Department of Computer Software Engineering, Soonchunhyang University, 22, Soonchunhyang-ro, Eumnae-ri, Asan-si,

Chungcheongnam-do, 35408, KOREA.
E-mail address: seokhoon@sch.ac.kr

1. 서론

지난 2016년 초 알파고 (AlphaGo)와 이세돌 9단의 세기의 대결로 인해, 국내에서 한동안 주목받지 못했던 인공지능 (AI: Artificial Intelligence)에 대한 연구개발이 다시 활발히 이루어지고 있을 뿐 아니라, 이와 관련된 다양한 신규 비즈니스가 창출되고 있는 실정이다. 이는 최근 활발히 연구개발이 이루어져 온 다양한 빅데이터 분석기법의 비약적 발전에 기인한 바가 크며, 이러한 추세 및 적용범위의 확대 역시 가속화 될 것이 확실시 되고 있다[1-4].

이러한 패러다임 변화는 정치·경제·사회·문화·과학 등 사회 각 분야에서 지속적으로 가속화되고 있는 실정이며, 이 중, MICE (Meeting, Incentive trip, Convention, Exhibition & Event) 비즈니스의 발전으로 대변되는 관광 산업과 빅데이터 분석 기반의 ICT 융합을 통한 고 부가가치 신규 비즈니스가 매우 각광받고 있다. 그러나, 관광 관련 비즈니스에서 빅데이터 기반 ICT 융합은 아직까지 단순 데이터 분석을 통한 Digital Marketing 촉진분야에만 주로 활용되고 있다[5]. 즉, 해당 비즈니스의 활성화 방안 마련을 위한 분석기법에만 초점을 맞추고 있을 뿐, 관광객을 위한 ICT 융합 서비스는 아직까지 매우 부족한 것이 사실이다. 때문에, 관광객을 위한 다양한 ICT 융합 서비스에 대한 연구개발이 시급히 필요한 시점이며, 특히 관광객들이 실질적으로 가장 필요로 하는 관광객을 위한 통·번역 시스템에 대한 연구개발이 시급히 수행되어야 함에도 불구하고 아직까지 이에 대한 연구개발은 거의 이루어지지 않고 있다.

한편, 다양한 빅데이터 분석기법의 발전은 자연어 처리 방안의 진화에도 큰 영향을 미치고 있으며, 특히 이러한 현상은 자연어 처리의 핵심 응용 분야 중 하나인 통·번역 시스템에서 두드러지게 나타나고 있다. 비록 심층 학습 (Deep Learning) 기

반의 빅데이터 처리방안의 급격한 발전으로 인해 이러한 통·번역 시스템의 정확도가 비약적으로 향상되었지만, 이는 주로 범용적인 구어체의 통·번역 결과에 대한 정확도 향상에 그치고 있으며, 아직까지 일상적인 회화체 위주의 통·번역 결과에서는 그 효과가 상대적으로 적은 것이 사실이다 [6]. 때문에, 주로 회화체 위주의 통·번역이 필요한 관광객을 위한 통·번역 시스템에서는 아직까지 많은 개선이 요구되고 있는 실정이다.

상기한 바와 같은 문제점들을 해결하기 위해 본 논문에서는 위치 기반 빅데이터 분석을 활용한 통·번역 시스템을 제안한다. 제안하는 통·번역 시스템은 기존의 빅데이터 분석방안을 활용하는 통·번역 시스템과는 달리 빅데이터 분석 시 사용자의 현재위치에 따른 빅데이터 분석을 기반으로 통·번역을 수행하는 시스템이며, 이를 통해 사용자는 각자의 상황에 맞는 통·번역 결과를 얻을 수 있을 뿐 아니라, 기존 통·번역기 대비 보다 현지화되고 자연스러운 구어체 통·번역 결과를 활용할 수 있다.

본 논문은 다음과 같이 구성되어 있다. 먼저 제 2 장에서는 통·번역 시스템을 위한 빅데이터 분석방안에 대해서 기술하고, 제 3 장에서는 제안하는 위치 기반 빅데이터 분석을 활용한 통번역 시스템 및 제안 시스템의 성능평가에 대해서 기술한다. 마지막으로 본 논문의 결론 및 향후 연구에 대해서는 제 4 장에 기술한다.

2. 빅데이터 분석방안

일반적으로 빅데이터 분석은 크게 다음과 같은 Customer Insight, Product & Process Efficiency, Digital Products & Service, Operational Excellence, Digital Marketing의 6개 카테고리에서 높은 효용 가치를 갖는 것으로 나타나고 있다[7]. 이러한 6개

카테고리의 효용 가치를 증대하기 위해 다양한 빅데이터 분석 알고리즘에 대한 연구개발이 꾸준히 수행되어 왔으며, 현재는 대부분 심층 학습 기반의 기계 학습에 대한 연구개발이 주를 이루고 있는 실정이다[8-10].

인공 신경망의 심층 학습 모델들이 좀 더 복잡하고 효과적인 모델을 구축할 수 있다는 장점이 있었지만, 기존 역전파 알고리즘 적용의 어려움 및 매우 높은 컴퓨팅 파워의 필요성으로 인해 제한적으로 이용되어 온 것이 사실이다. 하지만, 최근에는 Restricted Boltzmann Machine 기반의 사전 학습 (Pre-training) 기법의 적용 및 비약적 컴퓨팅 파워의 증대를 통해 이러한 문제들이 극복되었기 때문에, 이를 기반으로 한 다양한 방안들에 대한 연구개발이 활발히 수행되고 있다[11-13].

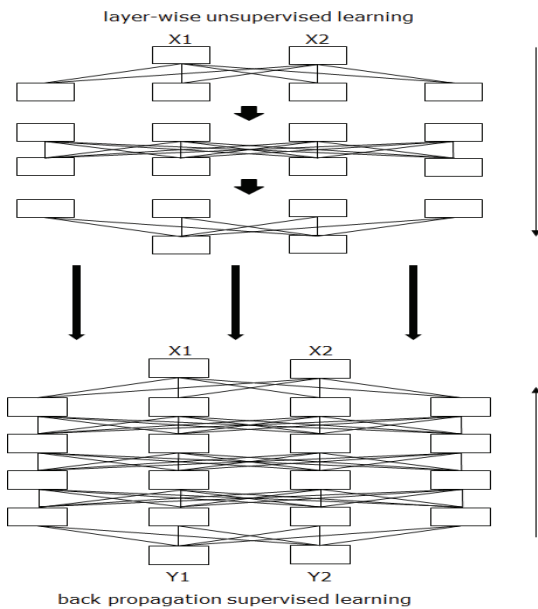


그림 1. 심층 신경망 학습을 위한 사전 학습
Figure 1. Pre-training for Deep Neural Network Learning

일반적으로 심층 신경망 학습에서는 사전 학습, 드롭 아웃, 조기 멈춤 등의 기법들이 사용되고 있

으며, <그림 1>에 도시한 기법은 본 논문에서 제안하는 시스템에서 사용되는 사전 학습이다. 사전 학습은 심층 신경망의 학습을 돕기 위해 사전에 정의된 중간 값을 미리 설정하는 방법으로써, 이를 통해 최적의 해를 찾을 확률을 높일 수 있다.

3. 위치 정보 기반 빅데이터 분석을 활용한 통·번역 시스템

제안하는 통·번역 시스템은 기본적으로 빅데이터 기반의 심층 학습을 바탕으로 통·번역 결과의 정확성을 향상시키는 것을 목적으로 하고 있으며, 이러한 측면에서는 기존의 통·번역 시스템과 큰 차이를 갖지 않는다. 그러나 제안하는 통·번역 시스템은 기존 통·번역 시스템과는 달리 심층 학습의 모태가 되는 심층 신경망 구축 시 사용되는 데이터를 수집할 때[14], 사용자의 위치를 기반으로 해당 위치에서 가까운 곳에서 생성된 데이터를 우선적으로 수집하여 심층 신경망을 구축한다는 점에서 가장 큰 차이점을 갖는다. 또한, 기존의 통·번역 시스템에서 제대로 지원하지 못했던 오디오 및 이미지에 대한 통·번역을 지원할 수 있다는 점에서도 커다란 차별성을 갖는다. 이와 같은 특성으로 인해, 제안 시스템은 사용자들이 실질적으로 필요한 구어체 위주의 자연스러운 통·번역 결과를 제공할 수 있으며, 이를 통해 자연스러운 대화를 이어갈 수 있도록 돕는다.

제안 시스템은 <그림 2>에 도시한 바와 같은 기법을 사용하여 통·번역을 수행한다. 이 때, 가장 중요한 프로세스는 통·번역의 수행을 위한 데이터 집합을 구축하는 것인데[15], 이러한 데이터 집합은 <그림 2>에 도시한 바와 같이 위치 정보를 기반으로 나이브 베이즈안 분류기에서 설정한 파라미터를 기준으로 가까운 곳에 위치한 데이터를 우선적으로 수집하여 생성한다. 이 후, 데이터를 수

집하는 로봇 에이전트가 해당 분류에 따른 이동을 계속하면서 데이터 집합을 생성하게 된다.

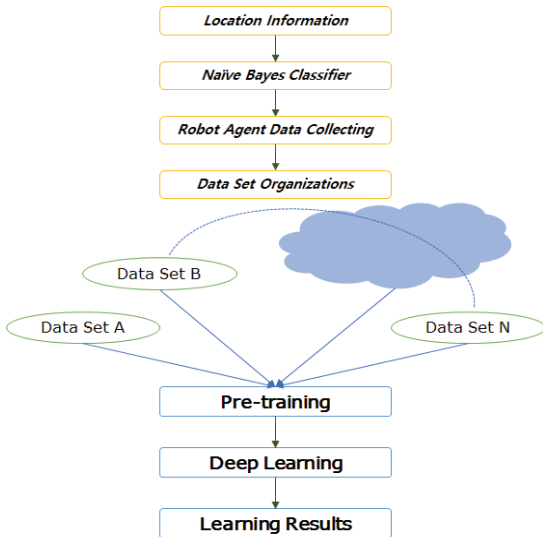


그림 2. 제안 시스템에서의 위치 정보 기반 사전 학습을 통한 심층 학습 기법

Figure 2. Deep Learning Method through a Pre-training based on Location Information in the Proposed System

본 논문에서 제안하는 시스템에서는 기본적으로 위치 정보에 기반한 나이브 베이저안 분류기를 통해 데이터 집합을 생성하는데, 이 때 데이터 집합을 생성하는 수식은 다음 <수식 1>과 같다.

$$\begin{aligned}
 DataSet &= \underset{b}{\operatorname{setmax}} P(b|a) = \underset{b}{\operatorname{setmax}} \frac{P(a|b)P(b)}{P(x)} \quad (1) \\
 &= \underset{b}{\operatorname{setmax}} P(a|b)P(b) \\
 &= \underset{b}{\operatorname{setmax}} \prod_{i=0}^{n-1} P(a_i|b)P(b)
 \end{aligned}$$

<수식 1>을 바탕으로 생성된 데이터 집합을 통해 사용자가 위치해 있는 지역에 대한 N개의 데이터 집합이 생성되며, 이는 이 후 수행될 심층학습의 중간값으로 사용된다. 이 때, 위치정보를 기반으로 데이터를 수집하는 로봇 에이전트는 사용자

의 위치를 나타내는 GPS 정보를 기반으로 반경 50Km 이내에서 생성된 데이터를 우선적으로 검색을 수행하나, 50Km 이내에서 수집된 데이터 정보만으로 통·번역에 필요한 데이터 집합의 절대 개수가 모자라는 경우에는 반경을 계속 넓혀 가며 충분한 데이터 집합의 생성이 완료될 때까지, 데이터 집합을 생성을 반복한다. 이러한 데이터 집합의 생성 과정은 텍스트, 오디오, 이미지 데이터의 수집에도 동일하게 적용된다. 또한, 텍스트, 오디오, 이미지에 대한 각각의 데이터 집합 생성 시, 모집단은 Google+, Facebook, Twitter가 되며, 이러한 사이트를 통해 수집된 데이터 집합이 심층학습의 중간값이 되며, 심층학습의 중간값 이외에 필요한 데이터 집합은 Google API를 활용했다. 결과적으로 이렇게 생성된 데이터 집합은 통·번역의 수행을 위한 사전 학습의 주요 파라미터 (사전 정의된 중간 값)로 사용되어, 통·번역 수행 결과의 정확도 향상에 기여하게 된다.

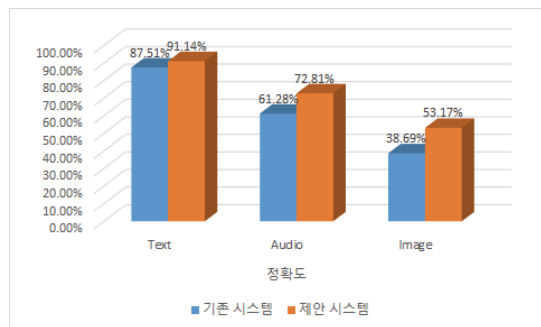


그림 3. 제안 시스템의 성능평가

Figure 3. Performance Evaluations with a Existing System

제안 시스템에 대한 성능평가를 위해 기존 시스템 (Real-time Interpreting and Translating System: RITS [16])과의 비교를 수행하였으며, 이 때 사용한 근거리의 범위는 도시와 국가를 기준으로 한다. 이를 통해 산출된 결과는 <그림 3>에 도시한 바와 같

다. 한편, 기존의 RITS는 주로 텍스트 통·번역에만 주안점을 둔 시스템이었으나, 본 논문에서 제안하는 시스템은 기존의 RITS와는 달리 오디오 및 이미지에 대한 데이터에 대한 번역도 제공할 수 있기 때문에, 기존 시스템 대비 통·번역 결과의 정확성을 향상시킬 수 있다. 제안 시스템의 성능평가는 크게 텍스트, 오디오, 이미지 데이터를 기준으로 수행되었으며, 각각 기존 시스템 대비 3.63%, 11.53%, 14.48% 향상된 것으로 나타났으며, 전체 시스템의 성능향상은 평균적으로 9.88%가 이루어진 것으로 나타났다.

기존 시스템 대비 제안 시스템의 성능이 향상된 가장 큰 이유는 통·번역 결과의 정확도 향상을 위해 수행하는 사전 학습에서 설정한 중간 값이 위치 정보를 기반으로 생성된 데이터 집합으로 설정되었기 때문인 것으로 판단된다. 이는 특정한 분류 없이 수집된 데이터를 통한 심층 학습 결과보다 사용자를 위해 좀 더 세밀하게 분류된 데이터 집합을 바탕으로 하는 심층 학습 결과가 정확도 향상에 기여할 수 있다는 것을 의미한다.

3. 결 론

본 논문에서 제안하는 통·번역을 위한 위치 기반 빅데이터 분석기법은 기본적으로 심층 신경망 학습을 위한 사전학습 시에 사용자들의 위치를 기반으로 수집된 데이터에 대한 사전학습을 수행하며, 이렇게 수행된 사전학습 결과를 심층 신경망 학습에 적용하여 학습결과를 산출하고, 이를 사용자를 위한 통·번역 시스템에 활용함으로써 보다 높은 정확도를 갖는 통·번역 결과를 얻을 수 있는 방안이다. 이를 통해, 제안하는 시스템을 사용하는 사용자는 현재 자신이 처한 상황에 걸맞고 보다 자연스러운 구어체 통·번역 결과를 활용할 수 있다. 그러나 아직까지 위치 기반 빅데이터 활

용을 위한 나이브 베이저안 분류기 적용 시 어떠한 파라미터를 적용했을 때 가장 적절한 분류가 될 것인지에 대한 보다 심층적인 연구가 필요할 뿐 아니라, 빅데이터 분석을 위한 데이터 수집 시 데이터 수집 에이전트의 적절한 이동방안 대한 보다 광범위한 향후 연구도 필요하다.

References

- [1] K. Panetta, *Top 10 strategic technology trends 2017, design and implementation of network virtual keyboard using smartphone*, Gartner, <http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/>, Oct. 2016.
- [2] N-J. Kim, and E-I. Choi, *Security technology framework in bigdata environment*, Journal of Knowledge Information Technology and Systems, Vol. 11, No. 3, pp. 251-260, Jun. 2016.
- [3] E. P. Xing, Q. Ho, W. Dai, J. K. Kim, J. Wei, S-H. Lee, X. Zheng, P. Xie, A. Kumar, and Y. Yu, *Petuum: A new platform for distributed machine learning on big data*, IEEE Transactions on Big Data, Vol. 1, Issue 2, pp. 49-67, Jun. 2015.
- [4] J-Y. Choi, J-H. Park, Y-G. Seo, S-W. Hong, and S-B. Kim, *A reliable service provided model for session hijacking attacks in big data service environments*, Journal of Knowledge Information Technology and Systems, Vol. 11, No. 6, pp. 597-606, Dec. 2016.
- [5] Korea Tourism Organization, *2015 Global online trend*, Research Report, Mar. 2016.
- [6] J. Dean, *Building Google: From search engine to AI poster child*, Structure Data 2016, Mar. 2016.

[7] J. K. Choi, *Current status and implications of big data analysis at domestic and abroad*, KISTEP InI, Vol. 2016, No. 14, pp. 33-43, 2016.

[8] Y-K. Ki, *Research trends in artificial intelligence and deep learning*, Weekly Technology Trends, Vol. 1736, pp. 10-19, Mar. 2016.

[9] O. G. Min, J. Y. Lim, and K. Park, *Research and development trends of artificial intelligence via evolving a bigdata*, Weekly Technology Trends, Vol. 1735, pp. 10-21, Mar. 2016.

[10] D. K. Kang, *Machine learning technology trends based on deep learning*, Weekly Technology Trends, Vol. 1742, pp. 12-24, Apr. 2016.

[11] N-J. Kwak, S-H. Park, and D-S. Kim, *Deep learning technology trends*, Weekly Technology Trends, Vol. 1745, pp. 13-25, May 2016.

[12] G. Hinton, Y. Benjjo, and Y. LeCun, *NIPS deep learning*, NIPS Tutorial, 2015.

[13] M. A. Alsheikh, S. Lin, D. Niyato, and H-P. Tan, *Machine learning in wireless sensor networks: algorithms, strategies, and applications*, IEEE Communications Surveys and Tutorials, Vol. 16, No. 4, pp. 1996-2018, Apr. 2014.

[14] A. Akusok, K.-M. Björk, Y. Miche, and A. Lendasse, *High-performance extreme learning machines: A complete toolbox for big data applications*, IEEE Access, Vol. 3, pp. 1011-1025, 2015.

[15] S. Nan, L. Sun, B. Chen, Z. Lin, and K-A. Toh, *Density-dependent quantized least squares support vector machine for large data sets*, IEEE Transactions on Neural

Networks and Learning Systems, Vol. 28, Issue 1, pp. 94-106, Jan. 2017.

[16] S-H. Kim, and S-J. Lee, *Real-time interpreting and translating system based on analyzing big data for tourists*, Indian Journal of Science and Technology, Vol. 8, No. 25, pp. 1-6, Oct. 2015.

위치 기반 빅데이터 분석을 활용한 통·번역 시스템

이상진¹, 김석훈²

¹창신대학교 호텔관광학과

²순천향대학교 컴퓨터소프트웨어공학과

요 약

최근 들어 심층학습 기반 빅데이터 처리방안의 급격한 발전으로 인해, 자연어 처리 기반의 통·번역 시스템의 정확성이 나날이 높아지고 있다. 그러나, 아직까지 회화체 위주의 통·번역이 필요한 사용자들이 사용하기에는 부족한 것이 사실이다. 이를 위해, 본 논문에서는 사용자 위치 기반 빅데이터 분석을 활용한 통·번역 시스템을 제안한다. 제안하는 통·번역 시스템은 데이터 수집 시에 위치 정보를 기반으로 해당 위치의 근처에 발생한 데이터를 우선적으로 수집하여, 이를 기반으로 데이터 집합을 생성하고, 실제 통·번역의 수행 시 해당 데이터의 집합을 심층 신경망 학습의 사전 학습 파라미터로 사용하여 구어체에 대한 통·번역 결과의 정확도를 향상시킬 수 있는 시스템이다. 제안 시스템의 우수성 검증을 위해, 기존 시스템과의 비교를 수행하였으며, 성능평가 결과 기존 시스템 대비 약 10%의 성능향상을 나타내어 제안 시스템의 우수성을 입증하였다.

감사의 글

본 연구는 순천향대학교 학술연구비 지원으로 수행하였음.



Sang-jin Lee received the bachelor's degree in the Department of Japanese from the Busan University of Foreign Studies in 1987. He received the M.S. degree in the Department of Japanese Literature Japan Chuo University in 1989. He received the Ph.D. degree in the Department of Japanese from the Busan University of Foreign Studies in 2007. He was a professor in the Dept. of Hospitality&Tourism at Changshin University since 1994. He was a Dean of Office of student affairs at Changshin University since 2013. He was a Mayor of human resources development center at Changshin University since 2015. He was a Mayor of Japanese Language and Literature Association of Korea-Japan from 2014 to 2015.

E-mail address: lee3137@cs.ac.kr



Seokoon Kim received the B.E. and Ph.D. degrees in computer engineering from Kyunghee University, Korea, in 2000 and 2004, respectively. From 2004 to 2006, he was with IPOne, Inc., Seoul, Korea, where he led various research projects as a Research Engineer. From 2006 to 2009, he was a Research Engineer at Neowave, Inc., Anyang, Korea, where he developed Mobile WiMAX (IEEE 802.16) devices. He was an Assistant Professor in the Department of Mobile Communications Engineering at Changshin University, Changwon, Korea. Since March 2016, he has been with the Department of Computer Software Engineering, Soonchunhyang University, Asan, Korea, where he is currently an Assistant Professor. His research interests

comprise Cloud Computing, Internet of Things, Software Defined Networking, Mobile System/Communications, and Machine Learning based on Bigdata.

E-mail address: seokhoon@sch.ac.kr