

펫 헬스 케어 서비스를 위한 GATs 기반 센서 데이터 처리 기법 설계

이지훈*, 문남미**

*호서대학교 컴퓨터공학과

**호서대학교 컴퓨터공학과

develomona@gmail.com, nammee.moon@gmail.com

Design of Sensor Data's Missing Value Handling Technique for Pet Healthcare Service based on Graph Attention Networks

Jihoon Lee*, Nammee Moon**

*Dept. of Computer Science, Hoseo University

**Dept. of Computer Science, Hoseo University

요 약

센서 데이터는 여러가지 원인으로 인해 데이터 결측치가 발생할 수 있으며, 결측치로 인한 데이터의 처리 방식에 따라 데이터 분석 결과가 다르게 해석될 수 있다. 이는 펫 헬스 케어 서비스에서 치명적인 문제로 연결될 수 있다. 따라서 본 논문에서는 펫 웨어러블 디바이스로부터 수집되는 다양한 센서 데이터의 결측치를 처리하기 위해 GATs(Graph Attention neTworks)와 LSTM(Long Short Term Memory)을 결합하여 활용한 데이터 결측치 처리 기법을 제안한다. 펫 웨어러블 디바이스의 센서 데이터가 서로 연관성을 가지고 있다는 점을 바탕으로 인접 노드의 Attention 수치와 Feature map을 도출한다. 이후 Prediction Layer를 통해 결측치의 Feature를 예측한다. 예측된 Feature를 기반으로 Decoding 과정과 함께 결측치 보간이 이루어진다. 제안된 기법은 모델의 변형을 통해 이상치 탐지에도 활용할 수 있을 것으로 기대한다.

1. 서론

최근 반려동물에게 인간이 누리는 서비스를 동일하게 제공하는 서비스가 증가하고 있다. 이는 반려동물을 인간처럼 대하는 펫 휴머니제이션(Pet Humanization) 현상의 가속화로 인한 추세이다. 이러한 이유로 사람에게 제공되는 IoT(Internet of Things) 기반의 웨어러블 디바이스를 통한 헬스 케어 서비스를 반려동물에게 제공하고자 펫 웨어러블 디바이스의 연구가 활발히 이루어지고 있다[1, 2].

웨어러블 디바이스로부터 수집되는 데이터는 센서로 수집되는 데이터가 주를 이룬다. 센서 데이터는 네트워크 통신 및 접촉 불량과 같은 다양한 이유로 인해 데이터 결측치가 발생할 수 있다. 인간은 웨어러블 디바이스의 결측치를 확인하면 직접 판단 및 해석이 가능하다. 하지만, 반려 동물에게 데이터의 결측은 왜곡된 데이터 해석으로 인해 치명적인 문제로 연결될 수 있다. 예를 들어, 가속도 센서를 통해 반려동물의 움직임을 파악하였다면, 반려동물의 온도가

높아지는 것은 지극히 정상적인 상황이다. 하지만, 가속도 센서로부터 받은 데이터가 정확하지 않아 움직임이 없었다면, 반려동물에게서 질병의 징후를 파악하지 못한 것이 된다. 본 논문에서는 이러한 왜곡된 데이터 해석을 방지하기 위해 GATs(Graph Attention Networks)와 LSTM(Long Short Term Memory)를 결합하여 활용한 데이터 결측치 처리 기법을 제안한다.

2. 관련연구

2-1. 시계열 데이터 결측치 처리

시계열 데이터에서의 결측치를 처리하기 위해 가장 널리 활용되는 방식으로는 데이터를 예측하여 보간하는 방식이다. 시간적 개념을 적용하기 위해 LSTM과 RNN (Recurrent Neural Network), 또는 GAN(Generative Adversarial Network) 등을 활용한 예측 방법이 많이 활용되고 있다[3, 4]. 이는 시간적 개념을 포함하는 것으로, 더 높은 정확도를 가진 예측 모델의 생성이 가

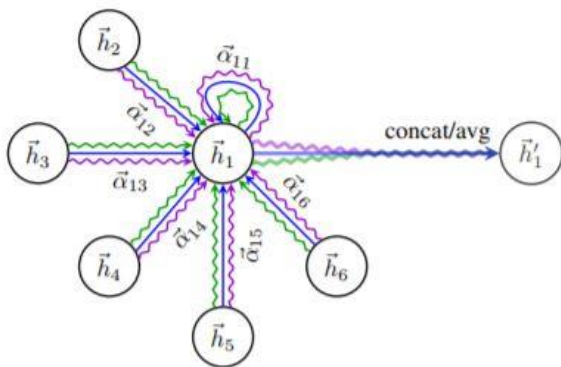
능하기 때문이다.

하지만 RNN 과 LSTM 을 활용하여 데이터의 시간적 특성을 반영하는 것으로 최대한 Gradient Vanishing 문제를 보완하였지만, Life-log 데이터와 같이 긴 시간에 걸쳐 측정되는 데이터에는 취약한 점이 있다. 이러한 문제점을 해결하기 위해 BiLSTM(Bi-Directional LSTM)을 활용하여 데이터의 시간적 특성을 양방향으로 반영하여 해결해왔다[5].

또 다른 방법으로는 Attention 기법을 활용하는 방법의 연구도 진행되고 있다. Attention 기법은 본래 NLP(Natural Language Processing)분야에서 RNN 의 Long-Term dependency 를 보완하기 위해 활용된 기법이다. 최근 이러한 특성을 활용하여 Attention 기법과 기존 시계열 데이터의 분석에 활용된 모델의 결합에 대한 연구가 다양하게 진행되고 있다[6].

2-2. Graph Attention Networks

GATs 는 기존 그래프 형태의 데이터를 학습하기 위한 네트워크 구조인 GCN(Graph Convolutional Network)에서 masked-self multi-head attention 을 도입한 것이다.



(그림 1) GATs 의 Architecture[7]

위 그림은 GATs 의 전체적인 구조를 보여준다. 결과적으로 중앙의 \vec{h}_1 은 주위의 다른 노드들에게 영향을 받아 \vec{h}_1' 로 진화된다. Attention 을 \vec{h}_2, \vec{h}_3 등 주위의 다른 노드들로부터만 받는 것이 아니라 \vec{h}_1 자신에게도 받으며, \vec{a}_{xy} 로 표현된 선들은 다른 노드들로부터 받는 Attention 을 표현한 것이다. Petar Veličković, et al.에 따르면, Convolution 을 Multi-channel 로 하는 것과 같이 Attention 도 Multi-head attention 을 사용한다. 이와 같은 GATs 의 특성은 교통 상황의 예측처럼 일련의 시간과 다양한 영향을 고려한 예측 기법에 활용되고 있다[8].

본 논문에서는 GATs 와 LSTM 을 결합한 모델을 통해 데이터를 보간하는 기법을 제안한다. 서로 연관성을 가지는 데이터에서 결측치가 발생할 경우, 제안된 기법은 보다 정확한 데이터를 예측하여 결측치를 보간할 수 있을 것으로 기대된다.

3. 센서 데이터 결측치 처리 기법

3-1. 펫 웨어러블 디바이스

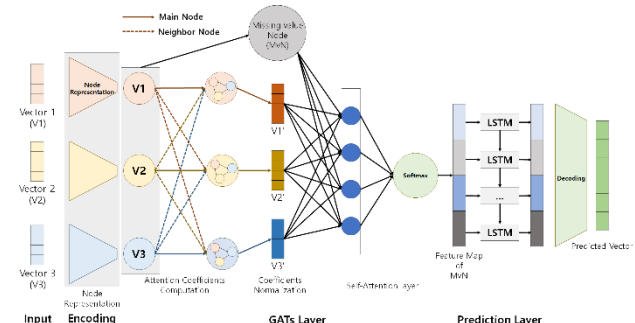
반려동물의 건강을 관리하기 위한 서비스는 다양하게 발전하고 있다. 그 중 원격 진료를 위한 반려동물 헬스케어 서비스의 반려동물용 웨어러블 디바이스도 다양하게 서비스되고 있다. 현재 서비스 중인 반려동물의 웨어러블 디바이스는 대부분 센서 데이터를 활용하여 운동량, 수면시간, 그리고 이상 징후의 탐지가 가능하다. 이는 진료용이 아닌 보호자에게 기본적인 모니터링 서비스를 제공하기 위함이다. 현재 반려동물용 웨어러블 디바이스는 기존에 서비스되는 단순 정보 전달에서 수의사의 전문지식을 통해 원격 진료가 가능하도록 발전하고 있다. 이런 동향에 맞추어 더욱 정확하고 정밀한 센서 데이터의 수집이 필요하다.

현재 서비스 중인 Petpace 는 체온 측정 및 활동량 측정, 그리고 심박수의 측정이 가능하다[9]. 하지만, Petpace 는 무거운 무게로 인해 소형 및 중형 반려동물이 많은 한국에서 원활한 사용이 어렵다. 따라서, 실험에 필요로 하는 웨어러블 디바이스를 제작하여 활용한다.

3-2. 제안 모델

본 논문에서는 반려동물로부터 웨어러블 디바이스를 통해 수집되는 데이터를 더욱 세밀하게 수집할 수 있도록 결측치를 보간하는 방법에 대해 제안한다. 반려동물로부터 수집되는 데이터는 최소 3 가지의 데이터가 필요로 하다. 각각의 데이터는 서로 영향을 주는 관계를 가진다. 예를 들어, 반려동물이 운동을 하고 있는 상황이라면, 가속도 센서를 통한 운동 데이터와 함께 심박 데이터는 상승하고, 체온 데이터 또한 일정 수치 이상 상승할 것이다.

이러한 상관관계를 가지는 데이터나 시계열적 특성을 가지는 데이터에서 결측치를 보간하기 위해 본 논문에서는 GATs 와 LSTM 을 통한 데이터 예측 기법을 제안한다.



(그림 2) 결측치 처리 기법의 전체적 흐름

위 (그림 2)는 본 논문에서 제안하는 결측치 처리 기법의 전체적 흐름에 대한 그림이다. 한 동물로부터 3 가지의 데이터를 수집하여 각각 Vector 1, Vector 2, Vector 3 으로 지정한다. 결측치를 보간하기 위해, 먼저

Node Representation 을 위한 Encoding 을 수행한다. 이는 각각의 데이터 간 Attention 수치를 구하기 위한 데이터의 노드화 작업이 필요로 하기 때문이다. 이후, GATs 를 통해 Attention Coefficients 를 산출하여 기준이 되는 노드와 인접 노드 간의 Coefficients 를 구한다. 이후 Coefficients 의 정규화 과정을 거치고 결측치가 있는 노드를 Self-Attention Layer 에 함께 입력 값으로 사용한다. 제안된 모델에서는 Softmax 활성화 함수를 사용하며 Feature map 의 산출을 위함이다. Softmax 는 입력 받은 값을 출력으로 모두 정규화하며, 출력 값들의 총합은 항상 1 이 되는 특성을 가졌다. 이를 이용하여 Feature map 을 산출하는 것으로, 필요로 하는 가중치를 구할 수 있다. 산출된 Feature map 은 데이터의 예측을 위해 LSTM 기반의 Prediction layer 를 거치게 된다. 예측된 데이터는 Vector 의 Decoding 과 함께 보간 작업이 이루어진다.

4. 결론

본 논문에서는 반려동물용 웨어러블 디바이스를 통한 센서 데이터의 더욱 세밀한 수집을 위해 GATs 와 LSTM 을 통한 결측치 보간 기법을 제안하였다. 제안된 기법에서의 State Attention Layer 의 활성화 함수는 추후 실험 과정에서 보완할 예정이다. 이는 더욱 정확한 결측치 예측을 위해 다양한 활성화 함수 중 최적의 함수를 찾을 필요가 있기 때문이다.

제안된 기법은 데이터의 전처리 과정에서 일어날 수 있는 데이터의 손실을 사전에 예방할 수 있을 것으로 기대한다. 또한, 기존 예측 모델을 활용한 데이터 보간 방식과는 다르게 서로 영향을 미치는 관계를 가진 요소들 간의 Attention Score 를 도출하여 보다 합리적이고 정확한 예측이 가능할 것이다. 더 나아가, 모델의 변형을 통해 기존 정상적인 센서 값을 학습하여 이상치 탐지에도 활용할 수 있을 것으로 기대한다.

사사

이 성과는 2021 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2012R1A2C2011966)

참고문헌

- [1] Roberto L. Junior, "IoT applications for monitoring companion animals: A systematic literature review", 2020 14th International Conference on Innovations in Information Technology (IIT), IEEE, 2020, pp. 239-246.
- [2] 김우찬, 김수균, 곽호영, "The Companion Animal Monitoring System using Low-Power Protocol Wearable Device", 한국컴퓨터정보학회논문지, vol. 25, no. 12, pp. 17-23, 2020
- [3] Qiuling Suo, Liuyi Yao, Guangxu Xun, Jianhui Sun, Aidong Zhang, "Recurrent Imputation for Multivariate Time Series with Missing Values", 2019 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2019, pp. 1-3.
- [4] Ying Zhang, Baohang Zhou, Xiangrui Cai, Wenya Guo, Xiaoke Ding, Xiaojie Yuan, "Missing value imputation in multivariate time series with end-to-end generative adversarial networks", Information Sciences, 551, pp.67-82, 2021.
- [5] 김진아, 문남미, "A deep bidirectional similarity learning model using dimensional reduction for multivariate time series clustering", Multimedia Tools and Applications, pp.1-3, 2021.
- [6] Richard Wu, Aoqian Zhang, Ihab F. Ilyas, Theodoros Rekatsinas, "Attention-based learning for missing data imputation in HoloClean", Proceedings of Machine Learnings and Systems, 2, pp.307-325, 2020.
- [7] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, "Graph Attention Networks", arXiv preprint arXiv:1710.10903, 2017.
- [8] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, Jianzhong Qi, "GMAN: A Graph Multi-Attention Network for Traffic Prediction", Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 01, pp. 1234-1241, 2020.
- [9] B. Belda, M. Enomoto, B.C. Case, B.D.X. Lascelles, "Initial evaluation of PetPace activity monitor", The Veterinary Journal, 237, pp. 63-68, 2018.