

라이프로그와 수면의 정보 연관성을 활용한 수면의 질 예측 알고리즘

An Algorithm for Sleep Quality Prediction by using Correlation of Lifelog and Sleep Information

요 약

최근 일의 다양성과 복잡성으로 인해 수면의 질이 저하되어 건강 문제 및 우울증이 증가하고 있다. 추가적으로 COVID-19 영향으로 인해 각자마다 생활패턴이 많은 변화를 이루게 되면서 수면의 질 향상이 필요하다. 이러한 문제를 해결하기 위해 BERT 알고리즘을 활용한 라이프로그와 수면 정보의 연관성을 추출하여 수면의 질을 예측하는 방법을 제안한다. 제안한 모델의 성능을 평가하기 위해 LSTM, CNN-LSTM을 베이스라인으로 설정하였고, 제안한 모델이 베이스라인 보다 더 잘 예측함을 보였다. 추가적으로 통계적 검증을 진행하여 유의미한 차이가 있음을 검증 하였다.

1. 서 론

최근 업무의 다양성과 복잡성으로 요구하는 일들이 많아짐에 따라, 수면의 질이 떨어지고 있다. 그리고 수면 장애가 심혈관 질환과 우울증 발병률에 영향이 있다는 연구 결과가 있어 수면의 질은 일상생활에 중요한 역할을 할 수 있다[1]. 또한 COVID-19로 인해 신체활동 저하, 수면 패턴 변화가 발생하였으며, 수면의 질 향상이 필요하다.

이와 관련하여 관련 연구로써, Chriskos et al.[2]은 피부온도, 심박수, 호흡수 등 다양한 생체신호를 수집하였다. 이러한 특징 정보를 활용하여 딥러닝 모델을 학습하고 최종적으로 수면의 질을 예측했다. Ezati et al.[3]은 규칙적인 유산소 운동이 수면의 질 향상에 영향이 있음을 보였다. Chang et al.[4]은 스스로 건강하다고 생각하거나 신체활동에 대한 만족도가 높을 경우, 수면의 질 향상에 좋은 영향이 있음을 보였다. 하지만 이러한 모델은 다양한 생체 정보와 데이터를 통해 모델을 검증해야 한다.

이와 같은 문제를 해결하기 위해 본 논문에서는 라이프로그와 수면의 정보 연관성을 활용하여 수면의 질 예측 알고리즘을 제안한다. 자세히는, 수면의 정보와 신체활동의 정보에 대한 연관성을 추출한다. 추출한 특징을 BERT 알고리즘 [5]의 입력으로 넣고 최종적으로 수면의 질을 예측한다. 제안한 모델의 성능을 평가하기 위해 베이스라인인 LSTM[6], CNN-LSTM[6]간 동일한 데이터에 대해 수면의 질 예측 성능을 비교를 하였으며, 제안한 모델이 수면의 질을 더 잘 예측하는 것을 보였다. 추가로 통계 검증[6]을 통하여 제안한 모델이 베이스라인보다 유의미한 차이가 있음을 검증하였다.

본 논문의 기여도는 다음과 같다.

- 라이프로그 정보와 수면의 정보 연관성을 추출하여 모델의 성능을 개선하였다. 각각 독립된 정보를 활용하였을 때보다 성능이 더 향상된 것을 보였다.

- 추출한 연관성을 활용하여 BERT 알고리즘에 학습을 진행하였고, 베이스라인보다 성능이 향상된 것을 보였다.

따라서 새로운 라이프로그 정보와 생체정보가 질의(Query) 된다면, 수면의 질을 예측할 수 있고 일상생활에 도움을 줄 수 있다.

2. 관련연구

딥러닝 알고리즘은 다양한 분야에 적용[5, 6]되고 있으며, 최근 수면의 질 예측 알고리즘에서 많이 활용되고 있다. 본 논문에 대한 관련 연구는 다음과 같다. Christin Lang et al.[7]은 신체활동-수면에 대한 연관성을 분석하였고, 수면의 질을 예측했다 가속도계로 PA(Personal Activity)를 측정하고 EEG(Electroencephalogram)로 수면을 측정했다. 연구의 결과로 PA가 높은 경우 수면시간이 더 길고 불면증 현상이 감소한 것을 보였다. Bisson et al.[8]은 하루 활동 시간에 대한 평균을 한 달 동안 측정했을 때 강도가 낮은 신체활동이 수면의 양과 관련이 있었으며 신체활동의 지속시간이 길수록 수면의 질이 향상함을 보였다. Aarti Sathyanarayana et al.[9]은 신체활동 정보 기반으로 HAR(Human Activity Recognition)을 수행하고 딥러닝을 적용하여 수면의 질 예측 모델을 제안했다.

3. 본 론

본 논문에서 수면의 질 예측을 위한 전반적인 도식도는 다음 그림 1과 같다. 먼저, 라이프로그의 데이터(신체활동 정보, 수면 정보)에 대해 전처리 과정을 진행한다. 그리고 신체활동 정보와 수면 정보에 대해 연관성을 추출한다. 추출된 연관성 특징을 활용하여 BERT 알고리즘 안에서 학습을 진행하여 그의 따른 결과값을 최종적으로 수면의 질에 대한 점수를 예측한다.

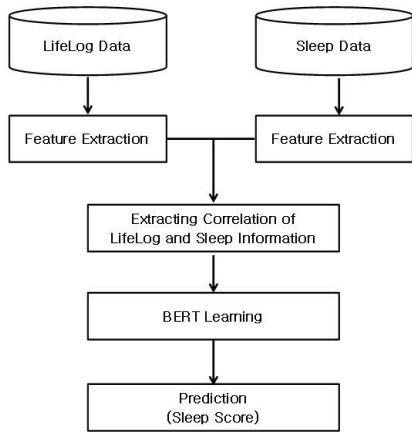


그림 1. 제안한 전반적인 도식도

3. 1 데이터 전처리

라이프로그와 수면의 특징을 추출하기 위해 데이터 전처리를 이용하여 정보를 추출한다. 라이프로그에는 신체정보의 키와 몸무게에 대해 체질량지수(BMI) 특징을 추출한다. 또한, 신체활동 분류 표[10]를 기반으로 활동 정보를 비활동, 좌식생활, 저강도, 중강도, 고강도로 분류한다. 추가적으로 TS(Timestamp) 기준으로 시간 단위로 환산하였으며, 여러 가지의 정보가 포함이 되었다면 최빈값을 표현한다. 예를 들어, 1시간 환산에서 저강도 활동이 3회, 비활동이 1회면 저강도 활동으로 분류한다.

3. 2 BERT 알고리즘

BERT 알고리즘은 사전 훈련된 딥러닝 모델[5]이다. 자세히는 Transformer 아키텍처를 사용하여 시계열 데이터의 순차적 정보를 인코딩하는데 효과적이고, 본 논문에서의 수면 시간 정보와 연관이 깊다. BERT의 구성은 사전학습 및 미세조정을 수행한다. 본 논문에서의 BERT 모델은 Transformer, Self-Attention, Feed-Forward Neural Network로 구성되며, Layer Normalization이 사용된다. 따라서, 추출한 라이프로그와 수면의 정보 연관성을 BERT 알고리즘의 입력으로 학습하고 최종적으로 수면의 질을 예측한다.

4. 실험

4.1. 데이터셋

제안한 모델을 실험하기 위해 ETRI 2020 라이프로그 벤치마크 데이터[11]를 사용한다. 자세히는, 사용자 22명의 라이프로그(행동), 수면 정보, 신체정보를 포함한다. 전체 데이터는 440,830개이며, TS를 시간 단위로 환산했을 때 9,151개이다.

4.2 평가척도

머신러닝 분야에서 자주 사용되는 Precision, Recall, F-Measure의 평가척도를 이용하고, 데이터의 편향(Bias)을 줄이기 위해 K-교차검증[6]을 진행(K=10)한다.

4.3 베이스라인

제안한 모델을 비교 평가하기 위해 다음의 베이스라인을 설정한다.

(1) LSTM: 장기 기억과 단기 기억을 가능하게 설계한 신경망의 구조의 모델로 주로 시계열 처리에 사용된다.

(2) CNN-LSTM: CNN에서 추출된 특징 정보를 LSTM에 입력으로 전달하여 시계열 데이터에서 공간적인 패턴을 학습할 수 있는 모델이다.

이 모델들은 시계열 분석 및 학습에 적합하고, 본 논문의 시계열 데이터와 적합하여 비교 대상으로 설정한다.

4.4 실험 결과

제안한 모델의 수면 질을 예측 결과는 다음 그림 2와 같다.

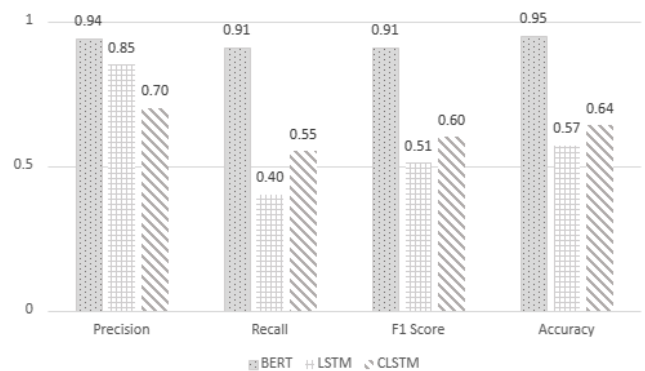


그림 2. 제안한 수면의 질 예측 성능 결과

자세히는, X축은 Precision, Recall, F-Measure를 의미하고, Y축은 결과에 대한 값(Value)을 나타낸다. 제안한 모델은 Precision에서 0.90, Recall에서 0.89, F-Measure에서 약 0.89, Accuracy에서 0.95의 성능을 보였다. 전반적으로 베이스라인 보다 수면의 질을 잘 예측함을 보인다. 추가적으로 라이프로그-수면 정보의 연관성 추출에 대한 성능을 다음 그림 3과 같이 보인다.

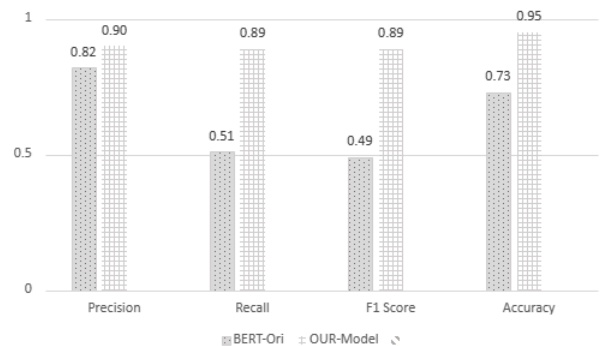


그림 3. 라이프로그 및 수면 연관성을 적용한 BERT(Our Model)과 적용안한 BERT 모델의 성능 비교

X축은 평가척도를 의미하고, Y축은 성능 값을 의미한다. Our Model은 라이프로그 및 수면 정보 연관성을 추출한

BERT 모델이고, BERT-Ori는 정보 연관성을 추출한 RAW 데이터 학습 모델을 의미한다. 제안한 정보 연관성을 추출한 모델이 더 좋은 성능을 보인다.

제안한 모델이 베이스라인보다 유의미한 차이가 있음을 보이기 위해 추가적으로 통계 검증[6]을 진행한다. 설정한 귀무가설(H_0)과 대립가설(H_a)은 다음과 같다

- H_{1_0} , H_{2_0} : 제안한 모델은 LSTM, CLSTM 보다 유의미한 차이가 없다는 것을 보인다.
- H_{1_a} , H_{2_a} : 제안한 모델은 LSTM, CLSTM 보다 유의미한 차이가 있다는 것을 보인다.

통계 검증을 진행하기 위해 먼저, 제안한 모델과, 베이스라인에 대해 정규성 검증[6]을 진행한다. 정규성 검증 결과 0.05보다 크면 T-검증[6]을 진행하고, 작으면 Wilcoxon-검증[6]을 진행한다.

다음 표 1과 같이 통계 검증 결과를 보인다.

표1. 통계 검증 결과

Null Hypothesis	p-value	Result
H_{1_0}	2.914E-07 (T-Test)	H_{1_a} : Accept
H_{2_0}	6.488E-05 (T-Test)	H_{2_a} : Accept

H_1 은 제안한 모델과 LSTM 알고리즘간 유의미한 차이가 없음을 나타낸 것을 의미한다. 통계 검증 결과 H_1 의 p-value는 2.914E-07(0.0000002914)이고 0.05보다 작으므로 귀무가설을 기각하고 대립가설을 승인한다. 따라서 제안한 모델은 LSTM 알고리즘 간에 유의미한 차이가 있음을 확인하고, 모든 귀무가설을 기각하고, 대립가설을 승인한다.

5. 연구결과

본 논문에서는 라이프로그 정보와 수면 정보의 연관성 추출을 활용하여 BERT 모델을 학습하였다. 베이스라인간 성능 비교에서 더 잘 수면의 질을 예측하였고, 통계 검증에서도 유의미한 차이가 있음을 보였다.

- 라이프로그 데이터와 수면의 정보 연관성을 추출하였고, 추출한 연관성으로 모델의 성능을 개선하였다. 베이스라인 보다 성능이 더 향상된 것을 확인하였다.
- 제안한 라이프로그-수면 정보 연관성을 적용한 모델이 RAW Data 학습한 BERT 모델보다 좋은 성능을 보였다.

6. 결 론

일상생활에서의 수면의 질은 생체리듬을 결정하는데 중요한 역할을 할 수 있다. 수면의 질을 예측하기 위해, 본

논문에서는 라이프로그 정보와 수면 정보의 연관성을 추출한다. 추출한 연관성을 BERT 알고리즘으로 학습하고, 최종적으로 수면의 질을 예측했다. 제안한 모델의 성능을 평가하기 위해 베이스라인 (LSTM, CNN-LSTM)과 비교하였으며, 제안한 모델이 더 잘 예측하였다. 추가적으로 통계 검증 또한 유의미한 차이가 있음을 보였다. 향후 다양한 생체정보와 특징 추출방법을 이용하여 모델을 더욱 구체화할 예정이다.

참고 문헌

[1] R. Robillard., K. Dion., M. H. Pennestri, E. Solomonova., E. Lee., M. Saad., A. Murkar., R. Godbout., J. D. Edwards., L. Quilty., A. R. Daros., R. Bhatla. & T. Kendzerska., "Profiles of Sleep Changes During The COVID-19 Pandemic: Demographic, Behavioural and Psychological Factors" Journal of sleep research, 30(1), 2021.

[2] P. Chriskos., C. A Frantzidis, C. M. Nday, P. T. Gkivogkli, P. D. Bamidis & C. Kourtidou-Papadeli., "A Review on Current Trends in Automatic Sleep Staging Through Bio-signal Recordings and Future Challenges" Sleep medicine reviews, 55, 101377, 2021.

[3] M. Ezati, M. Keshavarz , Z. A. Barandouzi & A. Montazeri., "The Effect of Regular Aerobic Exercise on Sleep Quality and Fatigue Among Female Student Dormitory Residents" BMC Sports Science, Medicine and Rehabilitation, 12, 1-8, 2020.

[4] S. P. Chang., K. S. Shih, C. P. Chi., C. M. Chang., K. L. Hwang. & Y. H. Chen., "Association Between Exercise Participation and Quality of Sleep and Life Among University Students in Taiwan" Asia Pacific Journal of Public Health, 28(4), 356-367, 2016.

[5] J. Devlin., M. W. Chang, K. Lee., & K. Toutanova, Jacob, Ming-Wei., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" arXiv preprint arXiv:1810.04805, 2018.

[6] C., Zhang, M., Cai, X., Zhao, and D., Wang, "Research on Case Preprocessing based on Deep Learning", In Journal of Concurrency Computat Pract Exper, vol. 34, no. 2, 2022.

[7] C. Lang., S. Brand., A. K. Feldmeth., E. Holsboer-Trachsler., U. Pühse. & M. Gerber., "Increased Self-reported and Objectively Assessed Physical Activity Predict Sleep Quality Among Adolescents" Physiology & behavior, 120, 46-53, 2013.

[8] A. N. xS. Bisson., S. A. Robinson. & M. E. Lachman., "Walk to A Better Night of Sleep: Testing The Relationship Between Physical Activity and Sleep" Sleep health, 5(5), 487-494, 2019.

[9] A. Sathyanarayana., S. Joty., L. Fernandez-Luque., F. Ofli., J. Srivastava., A. Elmagarmid., S. Taheri. & T. Arora., "Impact of Physical Activity on Sleep:A Deep Learning Based Exploration" arXiv preprint arXiv:1607.07034, 2016.

[10] E. K. Kim., H. Y. Jun., J. Y. Gwak. & J. O. Fenyi., K Eun-Kyung "Development of Physical Activity Classification Table For Koreans: Using The Compendium of Physical Activities in The United States" Journal of Nutrition and Health, 54(2), 129-138, 2021.

[11] ETRI KEMDy20 데이터 2023, "https://nanum.etri.re.kr/share/kjnoh/KEMDy20?lang=ko_KR", accessed April 30, 2023.