

트랜스포머 모델을 이용한 미래 혈당 예측 모델 개발

김서희^o, 김대연^{**}, 우지영^{*}

^o순천향대학교 ICT융합학과,

^{*}순천향대학교 ICT융합학과,

^{**}순천향대학교 천안병원 내분비내과

e-mail: {rlawjdgghl456, jywoo}@sch.ac.kr^o, c99851@schmc.ac.kr^{**}

Development of blood glucose prediction model using transformer model

Seohee Kim^o, DaeYeon Kim^{**}, JiYoung Woo^{*}

^oDept. of ICT Convergence, Soonchunhyang University,

^{*}Dept. of ICT Convergence, Soonchunhyang University,

^{**}Dept. of Endocrinology, Soonchunhyang University, Cheonan Hospital

● 요약 ●

본 연구에서는 순천향대학교 천안병원에서 제2형 당뇨병 입원 환자를 대상으로 연속 혈당 측정기(CGМ)를 통해 일주일 동안 수집된 101명의 혈당치 데이터를 사용하였다. 혈당치의 120분 동안 수집된 데이터를 기반으로 30분 후의 혈당치를 예측하는 트랜스포머 모델을 제안한다. 이는 트랜스포머의 인코더 모델만을 사용한 거보다 성능이 평균 제곱근 오차 (RMSE) 기준 약 4배 정도 향상하였으며, 이는 트랜스포머의 디코더 모델이 성능 향상에 효과적임을 보인다.

키워드: 트랜스포머(transformer), 혈당(blood glucose), 제2형 당뇨병(type 2 diabetes)

I. Introduction

대한당뇨학회에 따르면 2020년 기준 30세 이상 성인 6명 중 1명 (16.7%)이 당뇨병을 앓고 있으며, 이는 2018년 비해 2.9% 증가한 수치이다. 또한, 당화혈 색소가 8.0% 이상인 경우가 5명 중 1명으로 많은 환자들이 혈당 조절에 어려움을 겪고 있다 [1]. 당뇨병은 만성 합병증으로 이어질 수 있으므로 혈당 모니터링과 관리의 중요성은 점점 더 강조되고 있다. 최근 기술의 발전으로 연속혈당 측정기 (CGM)을 사용하여 당뇨병 환자들의 혈당을 더욱 쉽게 모니터링 할 수 있게 되었다. 이를 통해 대량의 환자 데이터 수집이 가능해졌으며, 이에 관한 연구가 증가하게 되었다.

본 연구에서는 시계열 데이터의 복잡한 장기 의존성을 포착할 수 있는 어텐션 메커니즘을 활용한 트랜스포머 모델을 사용하여 혈당 예측의 효과성을 검증하고자 한다.

II. Preliminaries

트랜스포머 모델을 활용한 혈당 예측에 관한 최근 연구에서 Lee et al. [2] 은 트랜스포머의 인코더만을 사용하여 예측을 수행하였다. 이는 120분 동안 수집된 혈당 데이터를 바탕으로 30분 후의 혈당 수치 예측에서 평균 절대 비율 오차 (MAPE) 16.82의 성능을 보였다. Sergazinov et al. [3]는 트랜스포머의 인코더와 디코더를 모두 사용하여 개인 맞춤형 혈당 예측 모델을 개발하였다. 이는 개인 맞춤형 모델로 새로운 환자에 대한 예측력 감소로 일반화된 결과 보장이 어렵다. 따라서 본 연구에서는 트랜스포머 모델의 인코더와 디코더를 모두 사용하는 동시에, 개인 맞춤형 모델이 아닌 전체 환자 데이터를 통합할 수 있는 모델을 개발하고자 한다. 이를 통해 혈당 예측 성능을 높이고, 모델의 일반화 가능성을 향상하는 것을 목표로 한다.

III. The Proposed Scheme

3.1 데이터

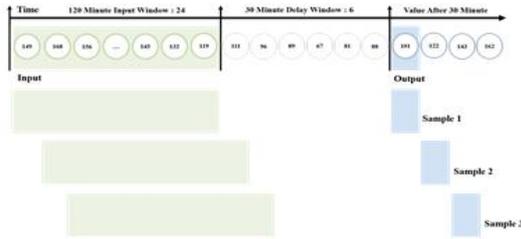


Fig. 1. Input Data Preprocessing Process

본 연구에서는 순천향대학교 천안병원의 제2형 당뇨병 입원 환자 101명으로부터 CGM 기기를 통해 7일 동안 수집된 데이터를 활용하였다 [4]. 실험에서는 입력 길이를 24개(120분)로 설정하고, 지연 시간은 6개(30분)로 하여 30분 후의 혈당값을 사전에 예측하도록 하였다. 이는 Fig. 1.과 같이 생성되었다.

3.2 모델 구성

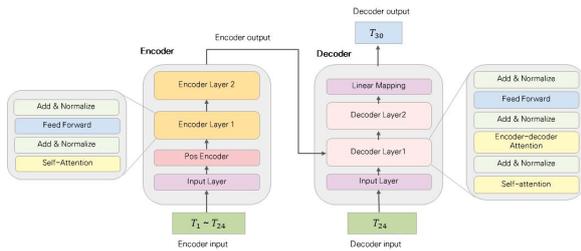


Fig. 2. Architecture of Transformer Model Based on Time Series Forecasting

제안하는 트랜스포머 [5]의 경우 Fig. 2.와 같이 구성하였다. 시간 단위 T_1 부터 T_{24} 까지의 데이터를 인코더의 입력으로 사용하였으며, 디코더의 입력으로는 인코더 입력의 마지막 값인 T_{24} 의 데이터만을 사용하였다. 최종적으로 6개의 지연 시간을 두고 T_{30} 의 혈당값을 예측하도록 설정하였다.

3.3 결과

Table 1. Forecasting result

Model	MAPE	RMSE
Transformer (Encoder & Decoder)	2.77	5.59
Transformer (Encoder)	11.75	23.22

본 연구에서는 트랜스포머 모델을 사용하여 실험을 진행하였다. 성능 비교를 위해 인코더만을 사용한 트랜스포머 모델을 사용하였다.

Table 1.에 제시된 결과에 따르면, 인코더만을 사용했을 때 성능은 MAPE 11.75, 평균 제곱근 오차 (RMSE) 23.22였던 반면, 인코더와 디코더를 모두 사용했을 때 MAPE 2.77, RMSE 5.59로, 성능이 약 4배 정도 향상되었다.

또한, 인코더만을 사용했을 경우 데이터의 양 끝단 예측에 어려움이 있었으나, 인코더와 디코더를 모두 사용한 실험에서는 이러한 문제가 상당히 개선됨을 확인하였다. 이는 인코더의 출력과 디코더의 입력 결합으로 예측 성능이 향상됨을 보여준다.

IV. Conclusions

본 연구에서는 제2형 당뇨병 환자의 미래 혈당 수치를 예측함으로써 사전에 혈당 이상 변화를 감지하는 것을 목표로 하였다. 연구 결과, 트랜스포머의 인코더와 디코더를 모두 사용하여 실험하였을 때 성능이 개선됨을 확인하였다. 또한, 현재 모델의 입력 길이는 120분으로 비교적 긴 시간이다. 향후 연구에서는 현재의 입력 길이를 단축하고 다양한 길이의 입력값을 사용한 실험을 통해 성능 비교를 할 것이다.

ACKNOWLEDGEMENT

본 연구는 한국연구재단의 지역대학우수과학자지원사업 (2020R111 A3056858)의 연구결과로 수행되었음

REFERENCES

- [1] Korean Diabetes Association, Diabetes Fact Sheet 2022, https://www.diabetes.or.kr/bbs/?code=fact_sheet
- [2] S.M Lee et al. Glucose transformer: Forecasting glucose level and events of hyperglycemia and hypoglycemia. IEEE Journal of Biomedical and Health Informatics, 27(3):1600–1611, 2023.
- [3] R. Sergazinov et al. Gluformer: Transformer- based personalized glucose forecasting with uncertainty quantification. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1–5, 2023.
- [4] D.Y Kim et al. Continuous glucose measurement for inpatient with type 2 diabetes. IEEE Dataport, 2023.
- [5] A. Vaswani et al., "Attention is all you need", Proc. Adv. Neural Inf. Process. Syst., pp.6000-6010, 2017.