ISSN 2636-0403

발표논문집

2025
Digital
Contents
Society

(사)한국디지털콘텐츠학회
하계종합학술대회 및 대학생논문경진대회

주최 | 사단법인 한국디지털콘텐츠학회 Digital Contents Society 주관 | 한신대학교 SW중심대학사업단, 목원대학교 MOKWON UNIVERSITY

후원 | JIN INFRA (주)신명프라, GNSOFT, (주)투비콤, KODICA 한국디지털컨버전스협회, INTECH DESIGN, 명화지리정보, 한국정보기술(주), M.I SOLTECH, 대보정보통신(주), (주)싸인텔레콤, elim 세림TSG(주), GFT 지에프티, SJ정보통신, ICTWAY 아이씨티웨이(주), CODABLE, JWVATS, OSCO, SIGONGtech, (주)하이제이컨설팅, OPENLINKSYSTEM, (주)에이치투테크, BonC Innovators, 대신정보통신(주), NEOBRIX, (주)그린텍아이엔씨, PINTEL, 지엔미래교육, D.PARTNERS, 힝어라

사단법인 한국디지털콘텐츠학회
Digital Contents Society

□ 논문발표 세션 WE3 (16:00~16:50)	○ 장소 : 카멜리아 홀
좌장	유선용 교수(전남대학교)

WE3-01. 제조기업 규모와 **RnD** 투자에 따른 디지털 전환 격차 분석
방수경, 김진술, 이명진(전남대학교) ····· 715

WE3-02. **YOLO v11** 기술을 이용한 실시간 스키드 탐지 모델 연구
김성권, 김진술(전남대학교) ····· 718

WE3-03. 연령별 안면 구분 및 맞춤형 키오스크 서비스
조혜원, 김수연, 조하영, 조영준(전남대학교) ····· 721

WE3-04. **Transformer** 기반 생물학적 그래프 모델을 활용한해석 가능한 약물 유도 유전자 발현 예측
송종웅, 서문수빈, 유선용(전남대학교) ····· 723

WE3-05. **Line-Crossing** 기반 과일봉지 계수 모델
윤경윤, 조영준(전남대학교) ····· 725

WE3-06. 분산신원증명을 이용한 티켓발원시스템 개선에 관한 연구
성기화(전남대학교), 안준영(㈜티맥스소프트), 김진술(전남대학교) ····· 727

WE3-07. **ChatGPT**를 활용한 디지털 콘텐츠 자동 생성의 가능성과 한계
이송배, 김진술(전남대학교) ····· 730

WE3-08. **Conditional Diffusion Model** 기반 약물로 인한 전사체 반응 예측
김채원, 정명현, 김민건, 유선용(전남대학교) ····· 732

WE3-09. 소규모 암시적 발화 데이터를 활용한 의도 분류 모델 성능 분석
이원주, 유연수, 김진술, 양형정(전남대학교) ····· 734

WE3-10. **C-ITS** 환경에서의 실시간 통신을 위한 **REST, gRPC, RSocket** 통신 프로토콜의 성능 비교 분석 및 적용 방안 연구
남현주, 옥상경, 김진술, 김경백(전남대학교) ····· 736

WE3-11. 디지털 지속가능성과 지속가능한 관광 간의 관계에서 인식된 경제적 이점과 친환경 인식의 조절 효과에 관한 연구
Barua Nishita, 박현재, Bithy Tahmida Amin(전남대학교) ····· 738

Conditional Diffusion Model 기반 약물로 인한 전사체 반응 예측

김채원(*), 정명현(*), 김민건(**), 유선용(***)
(*, ***) 전남대학교 지능전자컴퓨터공학과, {cwkim, mh_jeong, syyoo}@jnu.ac.kr
(**) 전남대학교 통계학과, 202720@jnu.ac.kr

Predicting Transcriptional Responses induced by Drugs based on Conditional Diffusion Model

Kim Chae-won(*), Jeong Myeong-hyeon(*), Kim Min-gun(**), Yoo Sun-yong(***)
(, ***) Chonnam National University, Department of Intelligent Electronics and Computer Engineering*
(**) Chonnam National University, Department of Statistics

요 약

본 논문에서는 Conditional Diffusion Model 기반 교란 조건을 고려한 전사체 변화 예측 심층 생성 모델을 소개한다. 처리한 화합물 정보와 더불어 처리용량과 시간, 세포주의 기저 유전자 발현 정보를 사용함으로써 정밀한 전사체 변화 예측을 가능하게 한다. 따라서 본 모델이 생성한 전사체 변화 데이터를 활용함으로써 약물에 대한 이해도를 향상하고, 신약 개발 및 정밀 의료 기술의 발전 등에 기여할 수 있는 가능성을 보여준다.

1. 서론

약물 교란으로 인한 전사체 변화를 이해하는 것은 생물학적 특성에 대해 근본적인 통찰을 제공하며, 신약 개발과 개인맞춤형 치료 전략 수립에 필수적이다. 그러나 모든 세포주-약물 조합의 전사체 변화 데이터를 실험적으로 얻는 것은 불가능하므로 최근에는 딥러닝 기반 교란 후 전사체 발현 값을 예측하려는 시도가 늘고 있다. 특히 Variantional Autoencoder (VAE) 구조를 기반으로 교란 후 전사체 발현 값의 분포를 예측한 연구는 세포주의 교란 전 전사체 발현 값과 화합물 정보를 입력으로 하여 훈련 과정에서 보이지 않은 (unseen) 화합물과 세포주에 대해서도 높은 예측 성능을 보였다[1]. 그러나 단일 샷 샘플링을 수행하기 때문에 조건에 따른 전사체 발현 변화의 다양성을 충분히 복원하기 어렵다는 한계가 존재한다.

따라서 본 논문에서는 Conditional Diffusion Model 기반 조건별 전사체 변화 예측 생성 기법을 제안한다.

2. 본론

2.1 Diffusion Model 기반 전사체 변화 예측

Diffusion Model[2]은 생성하고자 하는 데이터에 점진적으로 노이즈를 주입한 뒤, 이를 역으로 복원하는 과정을 통해 원본 데이터를 생성하는 딥러닝 기반 생성 모델이다. 이미지 생성 분야에서 그 성능이 입증된 이후 다양한 도메인에 응용되고 있으며, 전사체 발현 예측에서도 그 가능성이 주목받고 있다.

본 논문에서 제안하는 Conditional Diffusion Model은 세포주의 종류, 화합물, 처리농도, 처리시간 등의 교란 조건에 따라 교란된 전사체 발현 값을 복원하는 과정을 학습한다. 교란 조건이 유사하더라도 샘플링 경로에 따라 다양한 전사체 발현 양상을 생성함으로써 미세한 세포 반응 변화를 포착할 수 있기 때문에 VAE 구조의 한계인 단일 샷 샘플링에 따른 다양성 부족 문제를 해결할 수 있다. 본 연구에서 제안하는 모델의 구조는 그림 1.과 같다.

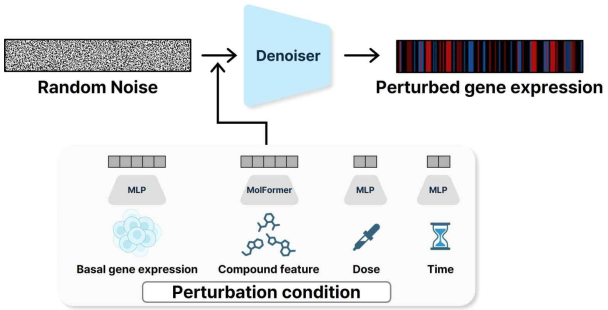

Figure 1. Perturbed gene expression predicting model architecture

그림 1. 교란 후 전사체 발현량 예측 모델 모식도

3. 실험 및 결과

3.1 데이터셋

본 연구에서는 LINCS L1000[3]에서 전사체 발현 데이터를 수집하였다. 전사체 발현 데이터는 각 교란 실험에 대한 세포주, 화합물, 처리용량, 처리시간 등의 정보를 제공한다. 전체 데이터에서 5번 미만 등장한 화합물은 제거하는 등의 전처리 과정을 거쳐 결과적으로 82개 세포주와 17,202개의 화합물에 대한 836,841개의 교란 후 전사체 데이터를 획득하였다. 이를 6:2:2 비율로 분할하여 각각 훈련, 검증 및 테스트 데이터셋으로 사용하였다.

3.2 결과

본 연구에서는 모델 학습에 Mean Squared Error (MSE)를 사용하였으며, 모델 성능 평가에는 Pearson Correlation Coefficients (PCC), Root MSE (RMSE) 를 사용하였다. 평가는 테스트 데이터셋의 전사체 변화 예측에 대한 성능을 기준으로 한다. 표 1.은 모델의 성능 평가 결과를 보여준다.

Table 1. Results of model performance evaluation

표 1. 모델의 성능 평가 결과

	Proposed model
PCC	0.57±0.05
RMSE	2.34±0.21

4. 결론 및 향후 계획

본 논문에서는 Conditional Diffusion Model을 기반으로 다양한 교란 조건 하에서 약물에 의해 변화한 전사체의 발현 값을 예측하는 모델을 제안하였다. 이를 통해 교란 조건에 의해 미세하게 변화하는 전사체 데이터를 생성할 수 있으며, 이는 전사체 데이터 부족 문제를 해결함으로써 신약 개발, 정밀 의료 기술의 발전에 기여할 수 있다.

향후에는 Latent Diffusion Model 및 Cross-Attention 등의 기법을 도입하여 모델의 성능을 개선하고, 기존 연구들과 비교 평가할 계획이다. 또한, 예측된 전사체 발현 결과에 대해 Gene Set Enrichment Analysis (GSEA)를 수행함으로써 모델의 생물학적 해석 가능성과 의의를 확보하고자 한다.

감사의 글

본 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원과 식품의약품안전처의 연구개발비로 수행되었음 (IITP-2025-RS-2022-00156287, RS-2024-00332003, RS-2025-02215961)

참고 문헌 (참고자료)

[1] Qi, Xiaoning, et al., "Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery," Nature Communications Vol. 15-1, pp.1~19, Octobor 2024.

[2] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in neural information processing systems, Vol. 33, pp.6840~6851, December 2020.

[3] Subramanian, Aravind, et al., "A next generation connectivity map: L1000 platform and the first 1,000,000 profiles," Cell, Vol. 171-6, pp.1437~1452, November 2017.