

[붙임] 결과보고서

[독립심화학습 결과보고서]

■ 과제명

과제명	심층 신경망을 이용한, XPS 데이터 분석	참여학기	2022 년 2 학기
-----	-------------------------	------	-------------

■ 강좌정보

과목명	독립심화학습 2	학수번호	PHYS331900
과제기간	2022 년 09 월 01 일 ~ 2022 년 12 월 20 일	학점	3

■ 팀구성

팀명	SpondLab Choi		팀구성 총인원	1 명
구분	성명	학번	학부(과)	학년
대표학생 참여학생	최정훈	2017103274	물리학과	4

■ 지도교수 확인

지도교수	성명	박용섭	직급	전임교원
	소속	물리학과	지도교수 확인	성명 : 박용섭 (인)

■ 붙임

[양식] 과제보고서

본인(또는 팀)은 과제를 성실히 이행하고 이에 따른 결과보고서를 제출합니다.

2022 년 12 월 21 일

제출자(또는 팀 대표): 최정훈  (인),

[독립심화학습 과제보고서]

과 제 명	국문	심층 신경망을 이용한 XPS 데이터 분석
	영문	Analysis of XPS data using deep neural networks
1. 과제 개요(선정 배경 및 취지, 내용 등)		
가. 과제 선정 배경 및 필요성		
X-선을 이용하여 물질 표면(Surface)의 화학적 구성 성분과 결합 상태를 분석할 수 있는 XPS(X-ray Photoelectron Spectroscopy)는 시료의 비 파괴 표면 분석이 가능하다는 장점으로 박막, 코팅, 촉매, 반도체 등 다양한 분야에서 사용되고 있다. 최근 주목 받고 있는 머신 러닝, 딥러닝 분야를 XPS 에 접목한 연구가 다양하게 진행되고 있다. 이를 이용하면 매우 짧은 시간안에 XPS 데이터를 분석해 낼 수 있을 것으로 기대된다.		
나. 과제 주요 내용		
임의의 양의 불순물이 포함된 물질에 대하여 측정된 XPS 데이터를 인공신경망을 이용해 불순물의 양과 원소 구성성분을 분석한다. 추가적으로 관측에서 사용되는 필터의 성질 때문에 악화될 수 밖에 없는 데이터의 해상도를 복구하는 방법과 상대적으로 적은(sparsity) 데이터를 복구하는 방법을 적용하여 본다.		
2. 과제 수행(연구) 방법		
가. 과제수행을 위한 연구방법		
기준에 발행되어 있는 논문을 토대로 연구를 수행한다. 물리학과 서버실에 있는 공용 서버를 이용해서 데이터를 생성 및 학습을 수행한다. 이를 통해 일정 수준(목표치)의 성능을 낼 수 있는 모델을 생성한다. Deep Neural Networks 모델을 학습시키기 위한 데이터 10k 개를 만들기 위한 방법을 찾는다. 해당 방법을 자동화 시켜서 목표 데이터를 생성한다. XPS 데이터를 분석하기 위한 모델을 학습시킨다. 위의 과정에서 적당한 파라미터를 알 수 없지만, Grid Search 를 통해서 파라미터를 얻기에는 학습 소요 시간이 클 것으로 예상되기 때문에 물리학적 지식을 기반으로 적당한 파라미터를 추론하고 이를 토대로 얻은 결과를 이용해서 비교한다. 적정 수준의 목표치(정확도)를 지정하고, 이를 달성하기 위해서 연구를 수행한다. 적절한 수준 이상의 모델이 생성되었을 경우, 이를 서버화 시킬 수 있는 방법에 대해서 찾고 이를 구현한다. 이후, Filter 와 광학적 특징으로 인해 발생하는 퍼짐 현상을 복구 시킬 수 있는 방법을 찾는다.		

3. 과제 수행(연구) 결과

연구는 크게 두가지 파트로 분할시켜서 진행되었다.

1) 데이터 생성(Generator)

미국 국립표준 기술 연구소(National Institute of Standards and Technology; NIST)에서 제공하는 XPS Database Sessa 를 이용해서 데이터 10k 개를 목표로 데이터 생성을 할 수 있는 Generator 시스템을 구축하였다. (이하 해당 시스템을 Generator Sys 라 칭함) Generator Sys 는 아래의 프로그램으로 구성되어 있다.

- generator: 메인 프로세서로 SESSA 와 직접적으로 통신하여 데이터 생성을 수행
- preprocessor: SESSA 를 통해 생성한 데이터를 읽고, DNN 모델에 학습하기 위한 형태로 변환 + SESSA 로그 파일 제거
- generator_master: generator & preprocess 오류 제어(Heart Beat 30sec 초과 시, 오류로 판단하고 reset 시킴)
- vanish_dir: Generator Sys 초기화(데이터 제거)

위 시스템을 통해서 시스템에 설정된 데이터(10k)에 도달할 때까지 안정적으로 데이터를 생성시킬 수 있도록 하였다. 이는 이 후 다른 X-ray Source, 측정 범위, random variable 에 대한 데이터를 생성하고 모델을 학습시키고자 할 때, 손 쉽게 수행할 수 있도록 하기 위함이다.

2) DNN Training

연구에서 사용된 DNN 의 모델은 논문에서 사용된 모델을 토대로 한다. Tensorflow 를 이용하여 이를 구현하였다. 논문에서는 전체적인 모델만 제시하였을 뿐, 파라미터를 정확히 명시되어 있지 않았기 때문에 몇 개의 파라미터를 이용해서 학습을 수행하였다. 최종적으로 학습을 끝낸 데이터는 검증 데이터(validation data) 및 시험 데이터(test data)에 대해서 높은 수준의 Contamination Depth 예측 능력을 가지게 되었지만, 특정 Substance Ratio 에 대해서 상대적으로 큰 오차를 출력하였다.

4. 결론 및 제언

연구를 통해서 얻은 모델은 입력한 Sessa's XPS 데이터에 대해서 학습이 되었다고 말할 수 있는 수준의 예측 능력은 보여주었다. 하지만, 높은 출력 데이터의 복잡성(80 개의 Atoms 에 대해서 Random 하게 5 개의 Ratio 로 구성)로 인해 해당 DNN 모델만 가지고는 더 높은 정확도를 얻는 것은 어려워 보였다. 더 높은 정확도를 얻기 위한 방법을 생각해 봤을 때, 아래와 같이 정리해 볼 수 있었다.

1. 물리적 지식을 토대로 실제로 영향력이 작은 원소 혹은 분자구조를 배제하고 학습을 수행

XPS 방법을 이용해서 원소를 측정할 때, Li 과 Be 같은 원소들은 타 원소들에 비해 낮은 영향력을 가지고 있다. 이는 모델로 검출하는 것이 불가능하며, 이들이 포함된 데이터를 예측했을 때, 큰 오차가 생기게 된다. 실제 물리상 검출이 힘든 데이터에 대해서 생성을 하지 않거나, 모델에 추가적인 가중치를 주어 이를 해결할 수 있을 것으로 보인다. 또한 우리가 알고 있는 화학구조 및 물리적 상태를 고려해서 분자구조의 종류를 제한해 줄 수 있으며, 이는 DNN 이 학습을 수행하는데 있어서 훨씬 유리한 환경을 제공해 줄 것이다.

2. 물리적 지식을 토대로 입력 데이터를 변환

입력데이터는 2048(0 to 1; float) 데이터로 구성되어 있다. 이를 하나의 입력데이터로 넣고 학습을 수행하였기 때문에 학습에 많은 데이터를 필요로 하며, 작은 Random 값에도 민감하게 반응하는 모델이 생성될 수 있다. 입력되는 데이터에 대해 추가적인 전 처리 과정을 통해 학습을 단순화 시켜줄 수 있을 것으로 기대된다. XPS 데이터에서 영향력이 높은 피크들의 Location & Intensity 를 추출하는 등의 물리적인 지식을 기반으로 데이터의 차원을 줄이거나, smoothing 방법을 전 처리 과정에서 수행해서 변화에 덜 민감한 모델을 만들 수 있을 것으로 기대한다.

독립 심화 학습 과제는 종료되었지만, 앞으로도 위에 방법들에 대한 추가적인 연구를 수행할 계획이다.

5. 참고문헌 및 조사기관

Deep neural network for x-ray photoelectron spectroscopy data analysis:

<https://iopscience.iop.org/article/10.1088/2632-2153/ab5da6>

NIST Database for the Simulation of Electron Spectra for Surface Analysis (SESSA): Version 2.2

<https://www.nist.gov/srd/nist-standard-reference-database-100>

※ 필요에 따라 양식 변경하여 사용 가능

제출자(또는 팀 대표) : 최정훈 (인) 