



The 19th International Conference on Intelligent Data Engineering and Automated Learning

2018.11.21-11.23

연세대학교 컴퓨터과학과
소프트컴퓨팅연구실
<http://sclab.yonsei.ac.kr>



발표 (1)

- 논문
 - 볼츠만머신을 이용한 가상화폐 거래용 강화학습 에이전트 선수학습 방법
- 발표 진행
 - 시간: 발표시간 10분, 질의응답 5분
 - 인원: 세션당 약 15명
- 결과
 - 언어가 다른 연구원들을 상당히 배려해주는 분위기
 - 영어공부 절실
 - 영어권 학생들에 비해 전달력 부족
 - 예상치 못한 질의응답 답변시의 어휘력 부족





발표 (2)

• 질의응답

- Q: 실제 온라인 거래소와 연결하여 투자실험한 결과인지

➔ A: 2016-2018년 데이터를 직접 수집하여 Back-test 수행함

- Q: 입력 데이터의 가로, 세로, 깊이축에 대한 설명 부탁

➔ A: 가로축은 시간으로 한칸 당 1분의 데이터를 나타냄. 정해진 길이만큼 샘플링하였고 실험에서는 45분 샘플링이 효과적이었음. 세로축은 가상화폐의 종류로 Bit, Eth,... 등등이고 깊이축은 8가지 서로다른 경제 지표임. 예를 들어서 시가, 종가, 고가, 저가 등

- Q: 가상화폐의 가격에 대한 최저와 최고를 예측할 수 없는만큼 0~1사이의 값으로 정규화가 어려울 것이다. 어떻게 정규화했나?

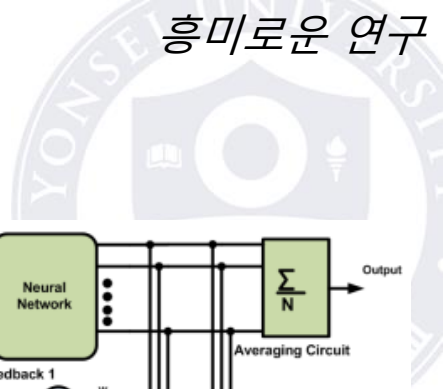
➔ A: 실제로 앞으로 얼마나 상승할 지 최고점을 모르는 상태이지만, 시뮬레이션에 기반한 Back-test 시에는 현재까지의 최저와 최고를 알고 있기 때문에 정규화할 수 있었고, 실제 온라인 상에서 투자실험하는 것은 앞으로 할 일중에 하나임

• 의견

- (좌장)주식시장을 포함하여 가상화폐 시장에서의 돈 흐름이 완전히 실시간이 아니므로, 추후 실험한다면 이부분에 대한 고려가 필요할 것

- 선수학습으로 DQN 네트워크의 입력공간을 축소한 흥미로운 케이스





Evolvable RNN 을 사용한 신호 재구축

- 22일 Evolutionary Computation 세션 발표논문
- 저자
 - Nadia Masood Khan and Gul Muhammad Khan
- 발표 내용
 - 문제
 - 시계열 신호 중 결측치의 대처방법
 - 시간 해상도가 낮은 시계열 데이터에 대처가능한 RNN
 - 방법
 - RNN의 구조 중 노드수를 인코딩하는 Chromosome
 - 진화연산을 이용하여 RNN 반복적으로 개선, Fitness = Accuracy
 - 결과
 - 결측치 25%, 50% 설정한 실험에서 타 방법 대비 27.39% 정확도 상승
 - 음성 데이터의 결측치 재구축시에 성능 검증
 - 비디오 데이터에 활용 가능성

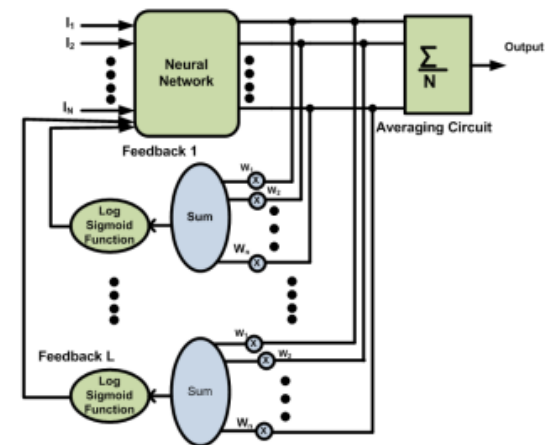
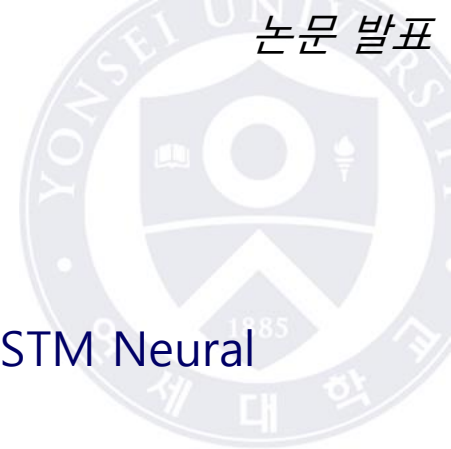


Fig. 2. A CGPRNN network with L feedback paths



발표 논문

- 논문 제목
 - Predicting Residential Energy Consumption using CNN-LSTM Neural Network

- 발표 진행
 - 시간: 발표시간 10분, 질의응답 5분
 - 셋째날, Session 12B
 - Deep learning & Neural network
 - 인원: 약 15명 내외

- 발표 결과
 - 질문이 활발하게 이루어지는 분위기
 - 청중들과 눈을 마주치며 발표하는 것이 중요
 - 영어 공부의 필요성
 - 질문에 대한 답변을 위해 시간 소모



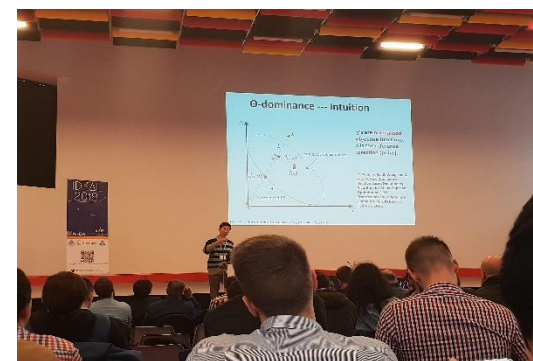


질의 응답

- Q1: 전력 수요 예측을 위해 계절 관련 변수를 사용하였나?
 - A1: 직접적인 계절 변수를 사용하지는 않았지만, month, day, hour 변수를 사용하여 간접적으로 계절 정보를 모델링 할 수 있도록 구현하였다.
- Q2: 학습된 모델의 입력과 출력 시계열 길이는 얼마인가?
 - A2: 입력, 출력 정보 모두 60분 단위로 전처리하였다. 단, 입력정보는 60분 단위의 7가지 변수를 포함하며 출력정보는 전력수요량이다.
- Q3: 60분 단위의 전력 수요를 예측하였는데 MSE 는 어떻게 계산하였나?
 - A3: 1분 단위로 비교한 MSE 의 평균값을 사용하여 계산하였다.
- 좌장 의견
 - Muti-step (60분 예측) 으로 전력 수요를 예측한 것이 강조되어야 함
 - 현재 Single-step (1분 예측) 과 헛갈릴 수 있음
 - Class activation map을 이용한 중요 변수 분석 실험이 흥미로움

Evolutionary Computation

- Detailed program
 - 둘째날 오전: Plenary talk, session 5A, 오후: session 8C
- Plenary talk
 - 연사: Xin Yao, 제목: tackling many objectives
 - 동적으로 진화하는 개체군을 통한 2차원 EA 구현
- Session
 - 총 7 편의 EC 논문 발표, 그 중 3 편 special session



Author	Method	Description
Michalak	Knowledge-based Solution	Graph구조와 진화연산을 이용한 시스템 위험 최소화 방법 제안
Michalak	Graph-based Solution	Firefighter Problem 해결을 위한 진화연산 방법 제안
Fister	Association rule Mining	Association rule를 최적화하기 위한 진화연산 제안
Islam	Feature selection	Rub-impact fault diagnosis를 위한 진화연산 기반 특징 추출
Suárez	Swarm optimization	Bat swarm optimization 기반 두 로봇의 목표지점 탐색 제안



발표 (1)

- 논문
 - 융합형 생성 모델을 통한 멀웨어 탐지
- 발표 진행
 - 시간: 발표시간 12분, 질의응답 3분
 - 인원: 세션당 약 15명
- 결과
 - 자유로운 토론 분위기
 - 발표가 끝나면 청중끼리 얘기도 나누고 질문을 대신 답해주기도 함
 - 영어 소통능력 한계
 - 다양한 영어 발음
 - 부족한 전달력





발표 (2)

• 질의응답

– Q: LSC의 의미가 무엇이고 왜 그렇게 정했는지

→ Latent semantic controlling 의 약자이고 명칭의 의미는 데이터를 특징에 따라 latent space에 전사함으로써 데이터의 의미적 성질을 조절할 수 있다는 것

– Q: 발표자료에서 재구축된 멀웨어 이미지는 퀄리티가 많이 떨어져 보이는데 학습에 도움이 되는가

→ 재구축하는 부분은 이미지의 특징을 추출하기 위한 LSC-VAE부분의 결과이며 멀웨어 이미지를 생성한 결과는 LSC-GAN을 이용하여 내었음

– Q: 멀웨어 코드와 이미지사이의 변환은 어떡하였는가

→ 코드에서 이미지로 변환하는 것은 이 슬라이드에 나타난 수식대로 16진수 코드의 가로/세로 길이를 변환하고 0~1로 스케일링하여 나타내었다.

• 의견

– 생성을 통해 지식을 넓히는 흥미로운 연구

– 생성된 멀웨어에 대한 분석이 있으면 더 좋을 것

Cloud 및 non-cloud data 사이의 전이학습을 통한 IDS

- 22일 Anomaly Detection 세션 발표논문

- 저자

- Roja Ahmadi, Robert D. Macredie, and Allan Tucker

- 발표 내용

- 문제

- 적은 수의 cloud intrusion 데이터셋
 - 많지만 다른 구조와 특징을 가진 non-cloud intrusion 데이터셋

- 방법

- Cloud 데이터와 non-cloud 데이터에서 비슷한 attribute들만 선택하여 데이터 구조 및 특징 일치
 - Non-cloud 데이터로 NB/DT/NN 학습 후 cloud 데이터로 검증

- 결과

- 다른 형태의 데이터셋으로 모델 적용 가능성
 - 학습 데이터와 검증 데이터셋 사이의 특징 비교를 통한 전이학습 성능 검증

Actual cloud attack and normal	Classified attack type and normal			The factor that the model suggests										
	Normal	TCP SYN	Smurf	Protocol_types			Src_bytes			Dst_bytes				
				TCP	UDP	ICMP	L	M	H	L	M	H		
Normal		✓		✓			✓							
TCP SYN	✓			✓	✓	✓	✓		✓	✓				✓
UDP flood	✓			✓	✓			✓			✓	✓	✓	
UDP flood		✓		✓			✓							
UDP flood			✓			✓								
DNS flood	✓				✓			✓	✓			✓	✓	
DNS flood		✓		✓			✓							
DNS flood			✓			✓								
Land	✓			✓	✓	✓	✓		✓	✓	✓			✓
Slowloris	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓		✓

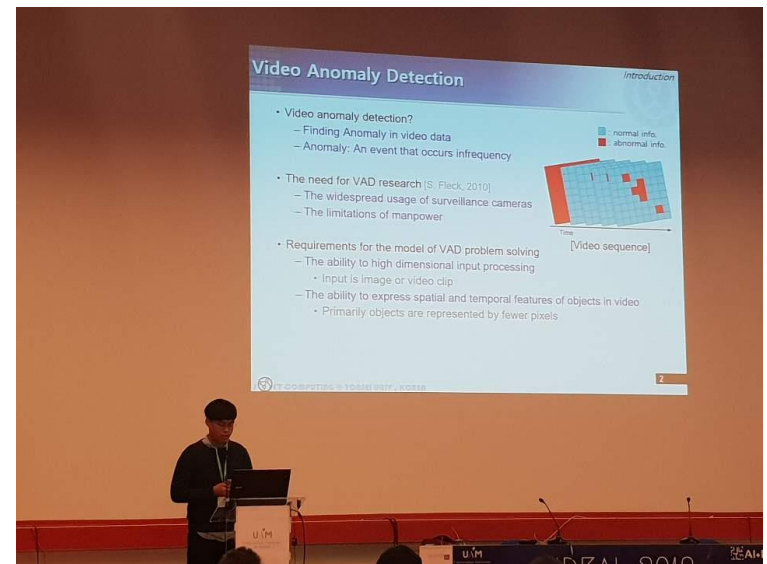


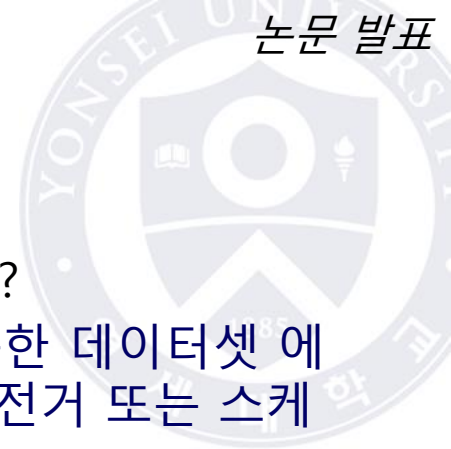
발표

- 논문
 - 영상이상검출을 위한 GAN기반 영상 시퀀스 생성 및 모델링 방법

- 발표 진행
 - 시간: 발표시간 12분, 질의응답 3분
 - 인원: 약 20~30명

- 결과
 - 청중들의 자유로운 표현과 제스처
 - 발표에 도움
 - 영어 회화공부 필요성
 - 다양한 발음 및 울리는 발표장





질의 및 응답

- Q1: 실제 영상에서 일어나는 이상의 종류는 어떤 것이 있는가?
 - ➔ A: 이상의 종류는 영상이 촬영된 장소 다름. 우리가 사용한 데이터셋 에서는 도보를 따라 걸어가는 사람들이 일반으로 분류되고 자전거 또는 스케이트보드를 타는 사람 및 자동차 등이 이상으로 분류됨
- Q2: 영상 블록을 입력으로 사용한 이유는 무엇인가?
 - ➔ A: 영상내 이상현상은 시간적 특성에 의해 발생할 수 있기 때문. 예를 들어, 한 장의 이미지에서는 달리는 사람과 걷는 사람을 구분 할 수 없음
- Q3: 영상 블록을 라벨링 한 근거가 무엇인가?
 - ➔ A: 영상내 이상현상은 한번 발생하면 수~수십 프레임이상 유지되는 특성이 존재. 따라서, 라벨링을 위한 영상 블록내 이상 프레임 수는 큰 문제가 아님
- 의견
 - 오탐지 케이스 분석을 통해 대부분 경우가 이상현상의 발생/종료 순간에 발생함을 밝혀낸 것은 흥미로움

안저 사진에서 삼출물 검출을 위한 지역 분석 전략

- 22일 image Analysis 세션 발표논문
- 저자
 - Joana Pereira, Adrian Colomer and Valery Naranjo
- 발표 내용
 - 문제
 - 당뇨병 망막증 조기 발견을 위한 망막 이미지에서의 삼출물 탐지 및 국소화
 - 방법
 - Waterpixel 방법을 사용한 국소화
 - LBP기반 특징 추출 및 SVM기반 이상탐지
 - 결과
 - 기존 방법대비 가장 높은 AUC

