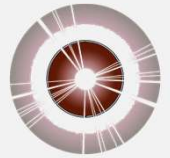




Variational Autoencoder를 이용한 경제 데이터 증강

최형규⁰¹ · 김래현² · 강재우^{2,3}

고려대학교 경영학과¹ ; 고려대학교 대학원 컴퓨터·전파통신공학과² ; 고려대학교 대학원 바이오협동과정³
{ imhgchoi, raehyun, kangj } @ korea . ac . kr



한국정보과학회
KOREAN INSTITUTE OF INFORMATION SCIENTISTS AND ENGINEERS

Abstract

Data has become key to the Industry 4.0 progress. Considerable amount of data is required for many recently proposed machine learning models. However, in certain domains such as the economic field, it may be difficult to attain sufficient samples of data. With the motivation to alleviate such deficiency, we propose a data augmentation methodology based on the Variational Autoencoder. By applying the method on the currency exchange rate data, we empirically show that the training model's performance and stability is improved when augmented data is added to the train dataset.

Motivation

- ① 데이터 증강[Data Augmentation]의 중요성
- ② 경제 분야의 데이터 부족 문제
ex. 환율 데이터, 주가 데이터

- ① Discrete Fourier Transformation (X. Teng *et al*, 2020)
- ② Self-Organizing Map (J. Zhang *et al*, 2017)
- ③ Variational Autoencoder?

Methodology

1 데이터 : 12개 국가[†] 환율

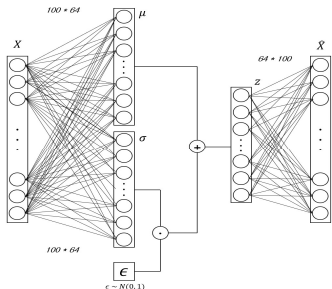
[†] Appendix 참조

Date	Open	High	Low	Close
2004-01-01	1191.5	1192.6	1181.5	1181.6
2004-01-02	1181.7	1195.8	1180.4	1195.8
2004-01-05	1190.4	1193.8	1178.6	1178.9
2004-01-06	1179	1189.7	1172.7	1189.4
2004-01-07	1189.3	1191.5	1173.3	1176.8
2004-01-08	1177.2	1183.2	1169.1	1171
2004-01-09	1170.9	1186	1157.7	1181.3
2004-01-12	1182.2	1183.9	1161.7	1168.8
2004-01-13	1168.5	1181.3	1166.2	1175.3
2004-01-14	1175.7	1195.1	1156	1175.2

Date	Open	High	Low	Close
2004-01-01	1.008464	1.009395	1	1.000085
2004-01-02	1.000169	1.012103	0.999069	1.012103
2004-01-05	1.007533	1.01041	0.997545	0.997799
2004-01-06	0.997884	1.00694	0.992552	1.006686
2004-01-07	1.006602	1.008464	0.99306	0.996022
2004-01-08	0.996361	1.001439	0.989505	0.991113
2004-01-09	0.991028	1.003809	0.979856	0.999831
2004-01-12	1.000592	1.002031	0.983242	0.989251
2004-01-13	0.988997	0.999831	0.98705	0.994752
2004-01-14	0.995091	1.011511	0.978417	0.994668

- 환율 [Foreign Currency Exchange Rate] 데이터 사용
- 학습 데이터 : 2004 ~ 2017 / 검증 데이터 : 2018 ~ 2019
- 국가 별 총 3,871 개 Sample (총 46,452개 Sample)
- 국가 환율 별 값 표준화
- 데이터 Sample 1개 : 100 days

2 데이터 증강 모델 : Variational Autoencoder



$$\log(p(x)) \geq \int \log(p(x|z))q_\theta(z|x)dz - \int \log \frac{q_\theta(z|x)}{p(z)}q_\theta(z|x)dz$$

$$= E_{q_\theta}[\log(p(x|z))] - KL(q_\theta(z|x) || p(z)) = ELBO(\theta)$$

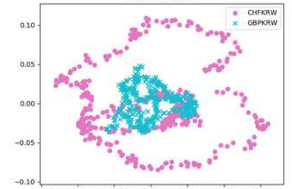
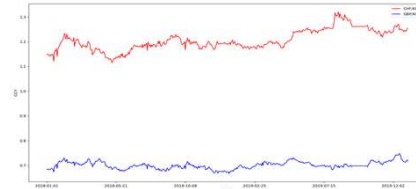
- ADAM optimizer 사용
- σ vector value clipping & μ vector Pre-Training 기법 적용

Why Variational Autoencoder?

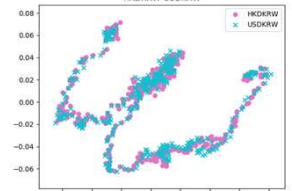
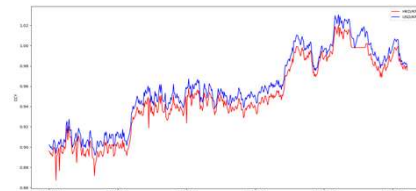
- 학습시키는 과정이 간단하고 직관적.
- 데이터 X의 우도(likelihood)를 추적할 수 있음.
- 생성하는 각 데이터의 확률밀도 함수값을 계산할 수 있음.
- 따라서 Data Augmentation Task에 적합.

Experiment Results

1 증강모델 학습결과 분석



(a) CHF/KRW와 GBP/KRW의 Latent Space Projection



(b) HKD/KRW와 USD/KRW의 Latent Space Projection

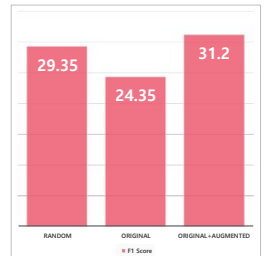
- 국가별 환율에 따른 잠재 공간(Latent Space)의 평균 벡터 μ 를 주성분 분석(Principal Component Analysis)을 통해 2차원 공간에 나타냄.
- (a)는 스위스 프랑/원 환율과 영국 파운드/원 환율의 잠재 공간을 도식화한 것. 서로 다른 환율은 서로 다른 잠재 공간에 맵핑됨.
- (b)는 홍콩 달러/원 환율과 미국 달러/원 환율의 잠재공간을 도식화한 것. 유사한 패턴을 가지는 환율은 유사한 잠재 공간에 맵핑됨.

2 증강데이터 실용성 검증

실용성 검증 실험 세팅

- 과거 환율 값들로 익일 환율 상승, 하강, 포함 여부를 예측하는 Multi-layer Perceptron 모델을 학습시킴.
(단, 포함은 |환율 변화율| > 0.01% 로 정함)
- 각 검증 데이터 샘플마다 100개의 데이터를 증강함.
- 다음 세 가지 방식으로 예측을 한 경우의 Accuracy와 F1 Score를 비교함.
 - 원본(Original) 데이터와 증강된(Augmented) 데이터를 모두 사용하여 학습시킨 MLP 모델로 예측.
 - 원본 데이터만을 가지고 학습시킨 MLP 모델로 예측.
 - 매 순간 무작위로 예측하는 랜덤 모델로 예측.
- 실험은 견고성 확보를 위해 10번 반복하여 진행.

(%)	Random		Original		Original+Augmented	
Iter.	Acc.	F1	Acc.	F1	Acc.	F1
1	34.14	30.14	44.43	20.51	46.99	31.14
2	34.12	30.38	48.42	21.84	47.02	31.29
3	32.69	29.29	48.29	22.85	46.95	31.18
4	32.87	29.04	48.20	21.68	46.82	31.16
5	32.71	29.00	49.01	33.66	46.93	31.11
6	32.67	28.95	48.24	21.70	46.86	31.13
7	32.62	28.86	48.31	21.72	47.10	31.36
8	33.81	29.72	48.46	21.76	47.02	31.24
9	32.54	29.06	47.87	31.51	46.56	30.89
10	32.69	29.07	48.11	26.31	47.35	31.52
Avg	33.06	29.35	47.93	24.35	46.96	31.20



- 증강 데이터를 사용한 경우 그렇지 않은 경우보다 평균 F1 Score가 6.85%p 가량 높음.
- 두 실험의 F1 Score에 대한 t-test p-value = 6E-4 < 유의수준 1%

Appendix

12개 국가 환율 목록

USD/KRW	Won/US dollar	CHF/KRW	Won/Swiss Franc
EUR/KRW	Won/European Euro	HKD/KRW	Won/HongKong Dollar
GBP/KRW	Won/England Pound	SGD/KRW	Won/Singapore Dollar
JPY/KRW	Won/Japan Yen	INR/KRW	Won/India Rupee
AUD/KRW	Won/Australia Dollar	THB/KRW	Won/Thailand Baht
CAD/KRW	Won/Canada Dollar	TWD/KRW	Won/Taiwan Dollar