2022 IEEE International Conference on Acoustics, Speech and Signal Processing

SENTIMENT-AWARE DISTILLATION FOR BITCOIN TREND FORECASTING UNDER PARTIAL OBSERVABILITY

Georgios Panagiotatos, Nikolaos Passalis, Avraam Tsantekidis, and Anastasios Tefas

E-mails: {panagiotga, passalis, avraamt, tefas}@csd.auth.gr

Computational Intelligence and Deep Learning Group (CIDL), AIIA Lab.

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece

DeepFinance

European Union
European Regional Development Fund

EPAnEK 2014-2020
OPERATIONAL PROGRAMME
COMPETITIVENESS
ENTREPRENEURSHIP
INNOVATION

ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

Co-funded by Greece and the European Union

ARISTOTLE UNIVERSITY OF THESSALONIKI

Structure

- **Introduction**
- **Proposed Method**
- **Experimental Evaluation**
- **Conclusions**

Introduction

- **Deep Learning (DL)** led to state-of-the-art results in numerous financial applications
- Most approaches rely on **price-related information** only
  - **e.g., Open-Low-High-Close candles**
- Human traders and analysts usually also take into account other information sources

Introduction

- **Sentiment** can be a valuable information source for various financial analysis tasks
  - e.g., sentiment expressed for cryptocurrencies in social media
- However, collecting this information and incorporating it into trading pipelines is difficult and costly
  - In some cases, it might be almost impossible to have up to date information (e.g., high frequency trading)

Contribution

- **Can we exploit the sentiment information that has already been collected to improve DL models for financial trading, while - at the same time - operating these models without access to such information?**

Proposed Method

- Simple, yet effective approach: **Employ sentiment information to co-supervise training**
- Implemented through **neural network distillation**
  - Train a model using **sentiment information**
  - **Distill the knowledge** from **the sentiment-aware model** into a **model that observes only price-related inputs**
    - **No need to have sentiment information available during inference!**

ARISTOTLE UNIVERSITY OF THESSALONIKI

icassp 2022

Input Features

- Price Features:

$$\mathbf{x}_t^{(p)} = \left[\frac{p_{t-L+1}}{p_{t-L+2}} - 1, \ldots, \frac{p_t}{p_{t-1}} - 1\right]^T \in \mathbb{R}^L,$$

where p_t denotes the close price of an asset at time t.

- Sentiment Features:

$$\mathbf{x}_t^{(s)} = [s_{t-L+1}, \ldots, s_t]^T \in \mathbb{R}^L,$$

where s_t provides the sentiment at the time step t.

Prediction Targets

- **Prediction targets:**

$$l_t = \begin{cases} 1, & \text{if } \frac{p_{t+1}}{p_t} - 1 > \delta \\ -1, & \text{if } \frac{p_{t+1}}{p_t} - 1 < \delta \\ 0, & \text{otherwise} \end{cases},$$

where δ denotes the threshold for considering a price movement significant

Neural Network Distillation

- **Teacher 1:** Trained on price to predict price movements
- **Teacher 2:** Trained on sentiment to predict price movements
- **Student:** Trained on price to predict price movements
  - **Distillation (i.e., training a student to mimic the teacher's output distribution mimicking through cross-entropy) from Teacher 1 and 2 simultaneously**
  - To enhance the **effectiveness of distillation**, we employ teacher ensembles instead of single teacher models for the distillation process.

Experimental Evaluation

- Experiments on **the Bitcoin-USD** currency pair
- Data collection period: from 2015 to 2020
  - First four years used for training, last year for backtesting

Experimental Evaluation

- **BDC Consulting dataset as sentiment source:**
  - Contains over 200,000 titles of financial articles that have been collected from various online sites
  - Used a pre-trained **FinBERT** model for sentiment extraction
  - Sentiment encoded from -1 (negative) to 1 (positive)

Model Architectures

- Long Short Term Memory (LSTM)-based architecture for models with price input
- MLP-based architecture for models with sentiment input
- Adam was used for the optimization, while models were trained for 100 epochs
- Ensemble: 5 teachers

Experimental Evaluation

- Profit and Loss (PnL)-based evaluation
  - 10 repeated runs for all experiments
  - Baseline experiments: trained on price or sentiment modality
  - Cross-distillation experiments: transferring from the oppositive modality

Experimental Evaluation

Method / Epoch	50	70	100
Price - Baseline	0.198	-0.233	-0.398
Sentiment - Baseline	0.148	-0.348	-0.235
Price - Cross Distillation	0.946	0.483	0.048
Sentiment - Cross Distillation	1.246	1.002	0.923
Price - Proposed	**1.334**	**1.354**	**1.155**

Conclusions

- **Presented a method that exploits sentiment information as a source of additional supervision during the training process, improving the accuracy of the developed models**
- **Future research directions:**
  - Distillation using other modalities, e.g., news articles, forecasts, etc.,
  - Other ways of employing sentiment to supervise the training process

ARISTOTLE UNIVERSITY OF THESSALONIKI

Acknowledgements

This work has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE (project code: T2EDK-02094)

Find more about DeepFinance project at deepfinance.csd.auth.gr

DeepFinance

European Union
European Regional Development Fund

EPAnEK 2014-2020
OPERATIONAL PROGRAMME
COMPETITIVENESS
ENTREPRENEURSHIP
INNOVATION

ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

Co-funded by Greece and the European Union

Thank you!