2022 International Conference on INnovations in Intelligent SysTems and Applications

Online Knowledge Distillation for Financial Timeseries Forecasting

Pavlos Floratos, Avraam Tsantekidis, Nikolaos Passalis and Anastasios Tefas

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

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

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

DeepFinance

ARISTOTLE UNIVERSITY OF THESSALONIKI

Introduction

- Introduction
- Proposed Method
- Experimental Evaluation
- Conclusions

Introduction

- **Deep Learning (DL)** led **to state-of-the-art performance** for several tasks, including **financial time-series analysis**.

- Several DL-based approaches have been proposed to this end.
    - **Classification-based setups typically prevail**.
    - Other approaches do exist, e.g., Deep Reinforcement Learning.

Introduction

- **Deep Neural Network (DNN) classifiers** can be used to this end in several different settings.
- Such networks operate on various **information regarding a time-series**.
    - e.g., past daily returns of an asset, volatility, etc.
- Typical setup: **predict the return over a specific horizon**, e.g., daily returns.

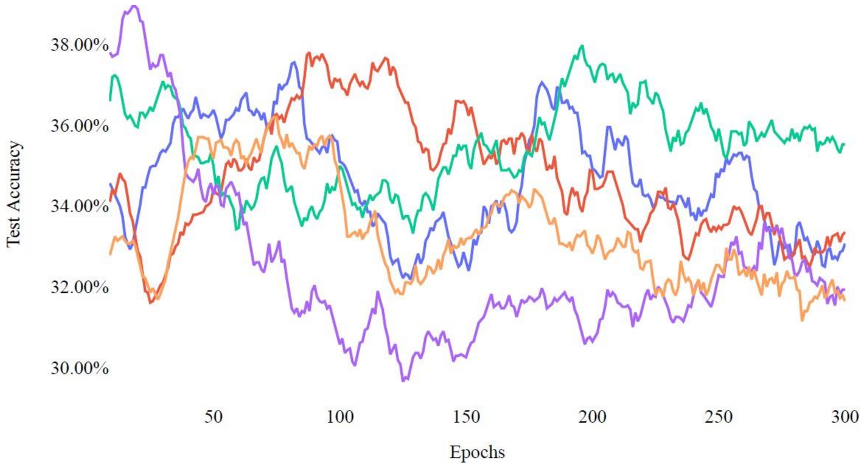
Introduction

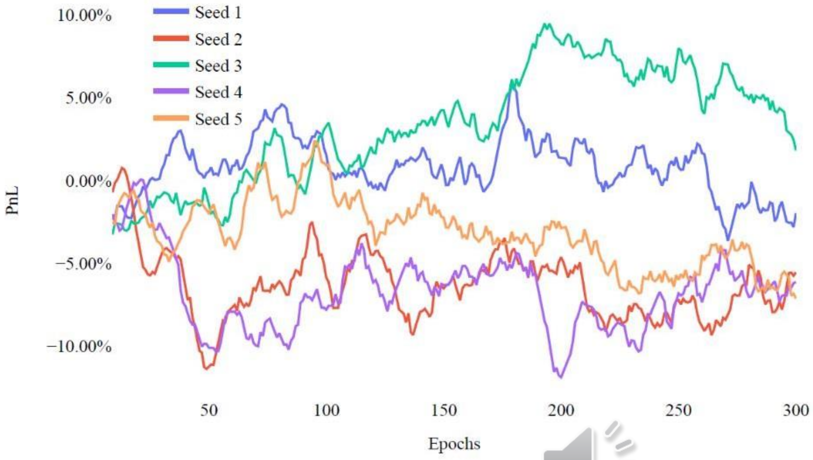
- However, **employing DNNs on financial time-series analysis tasks** is a **notoriously difficult and unstable problem**.
- The **noisy nature** of the data leads to **considerably different behaviours between models,** even when trained with **exactly the same hyper-parameters**.
    - **Affects a wide range of settings and setups**, e.g., both classification and reinforcement learning-based.

ARISTOTLE UNIVERSITY OF THESSALONIKI

Introduction

Comparing five different training runs using different initializations of the same deep residual CNN.





Introduction

- The classification set-up always **introduces an additional hyper-parameter**.
    - **Threshold for price movements** to be assigned to each of the **available classes (e.g.,** up, down, stationary).
- The aforementioned instability issues could be (partially) attributed to the existence of near-conflicting ground truth annotations.
- **When similar input samples carry different annotations, the network is forced to overfit the noise component of the data**.

Introduction

- Aim: **Mitigate the intense overfitting problem caused by conflicting annotations in financial trading agents**.

- Contribution:
  - Employed **ensemble-based online knowledge distillation** to transform the hard ground truth labels into soft and more meaningful representations.
  - Stabilize the training process and the performance of the networks.

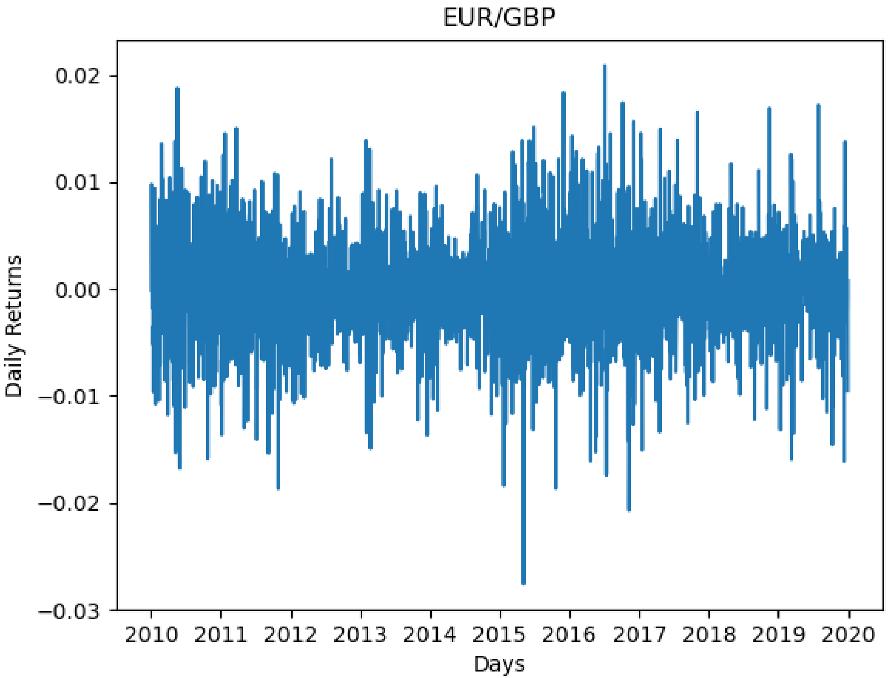
ARISTOTLE
UNIVERSITY OF
THESSALONIKI

Proposed Method

ARISTOTLE UNIVERSITY OF THESSALONIKI

FOREX Dataset Structure

- Features: $[R_{t-w+1}, \ldots, R_t]$
- Labels: [Buy, Sell, Exit]
  Classes' split based on R_{t+1}. All classes are perfectly balanced in train set



ARISTOTLE UNIVERSITY OF THESSALONIKI

Proposed Method

- **Knowledge distillation** can be used to transform the **handcrafted labels** to **more meaningful representations**.
- **More information regarding the similarity of samples with the available classes is introduced**.

Basic Knowledge distillation set-up:

- **Train 1 teacher** network for some epochs with the handcrafted labels.
- **Use the teacher network to produce soft representations** that will be used for the **student's training**.

Proposed Method

Drawbacks

- **The teacher can be unreliable** as it was trained with conflicting hard representations.
- **Computational intensity** as this is a **two-step process that cannot be parallelized**.

Proposed Method

<u>Ensemble Based Knowledge Distillation</u>

Steps:

1. A number (e.g., 5) teacher networks are trained for N epochs (e.g., 10) with hard labels.
2. The training stops and labels are extracted for all data from the teachers ensemble.
3. One student network is trained with soft labels coming from the ensemble.

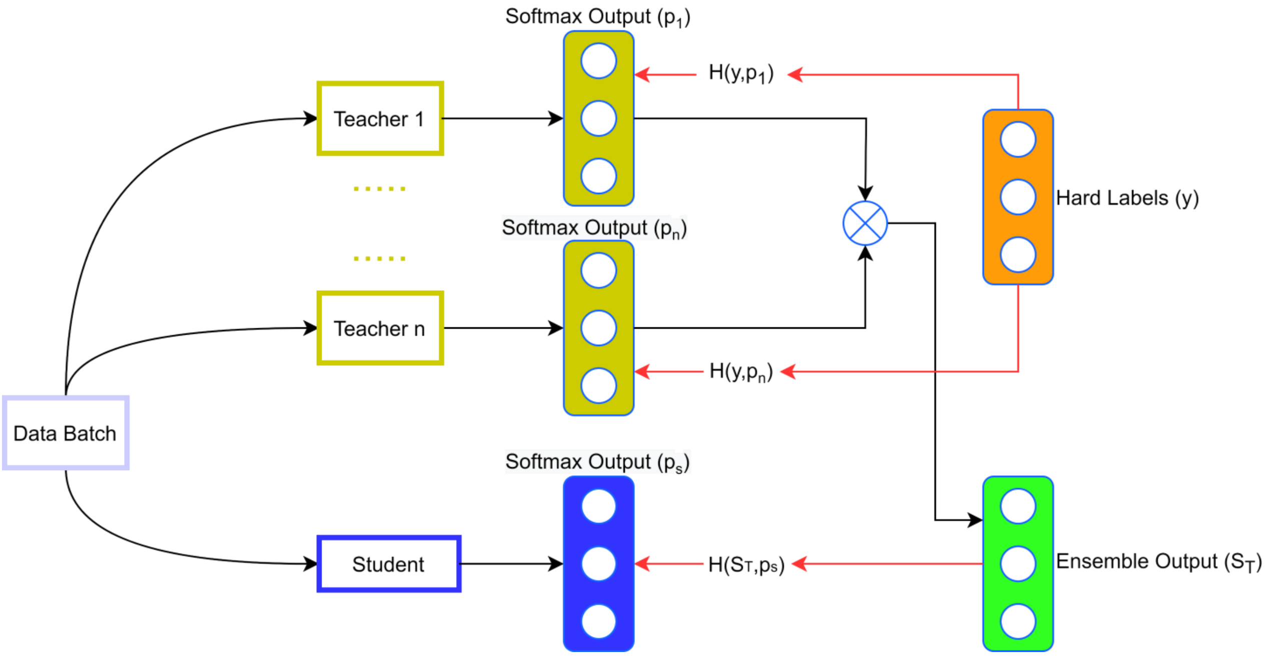
Proposed Method

- Developed a Deep Learning financial trading agent via **Ensemble Based Online Knowledge Distillation.**

- Teacher networks are used **to gradually transform the initial hard ground truth labels** into more **meaningful and less conflicting representation.**

Proposed Method

Ensemble Based Online Knowledge Distillation



Advantages:

- Easy to use one step end-to-end process.
- Reduces computational complexity as teachers and student can be trained in parallel.

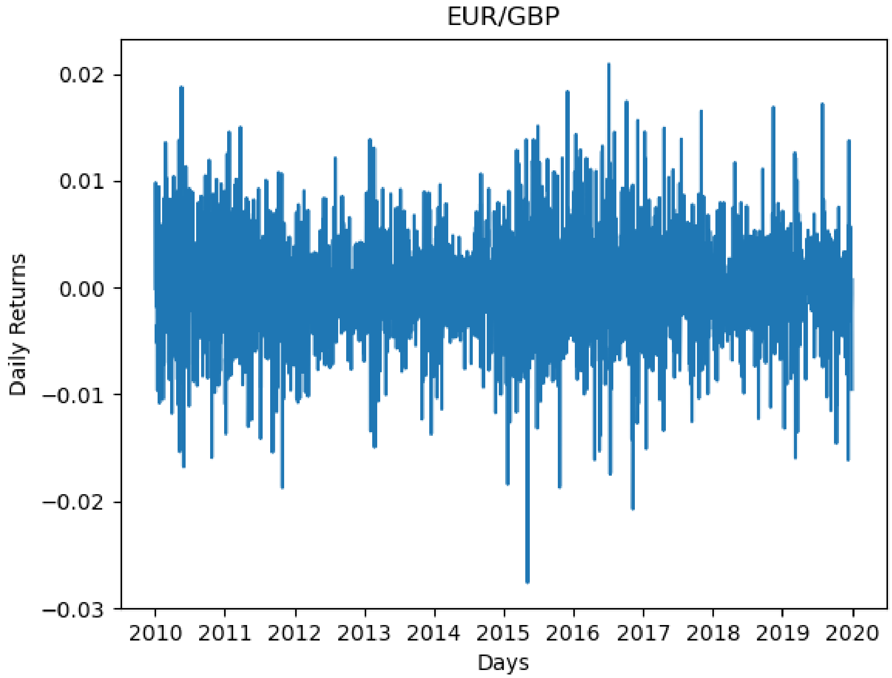
ARISTOTLE UNIVERSITY OF THESSALONIKI

Experimental Evaluation

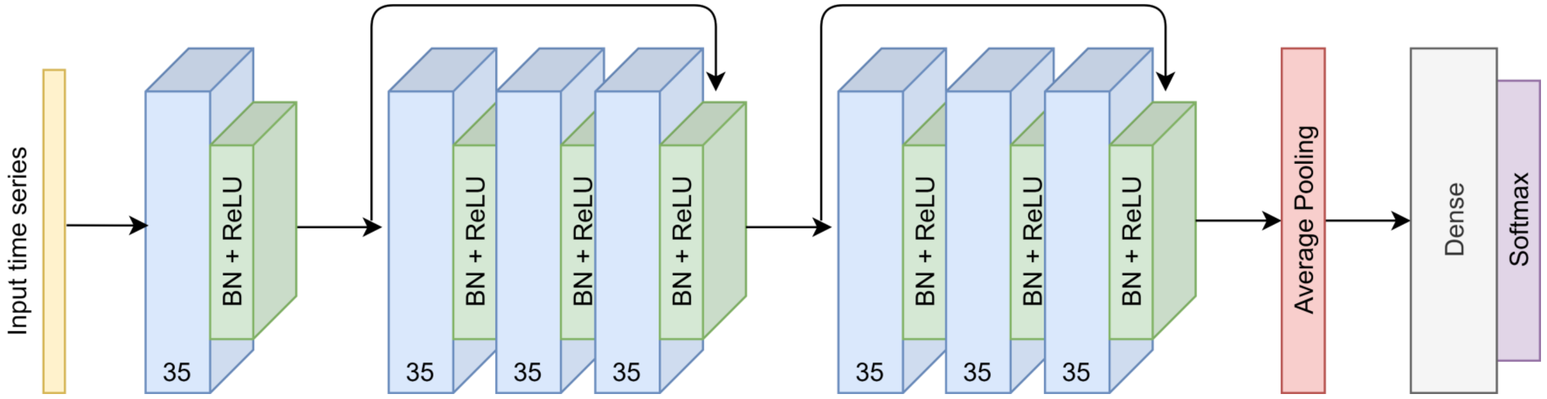
ARISTOTLE
UNIVERSITY OF
THESSALONIKI

FOREX Dataset

37 different currency pairs were used such as *EUR/GBP, EUR/USD, CHF/JPY, GBP/CAD, USD/NOK* with a total of 114,234 samples.



Model Architecture



On all following experiments the exact same network architecture and hyper-parameters were used, with the only difference being in the applied knowledge distillation method.

Experimental Set-Up: Walk Forward Testing

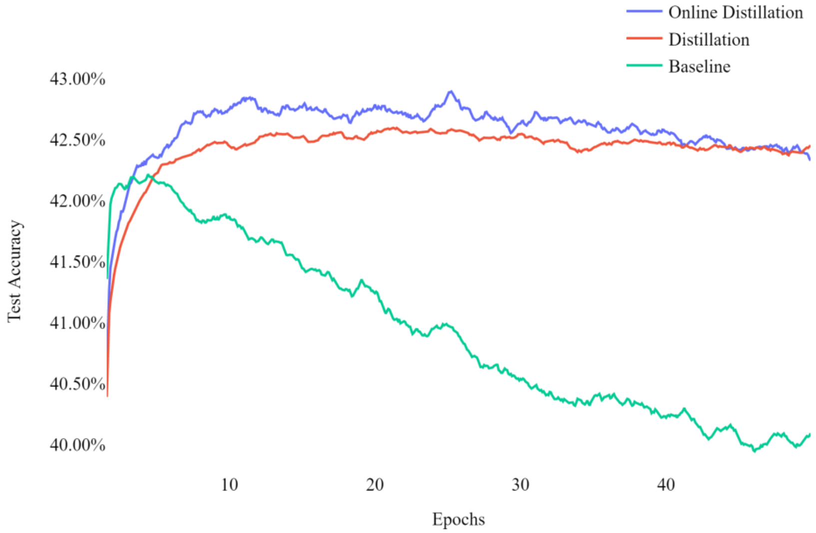
<table>
<tr><th>Experiments/ Train-Test</th><th>2010</th><th>2011</th><th>2012</th><th>2013</th><th>2014</th><th>2015</th><th>2016</th><th>2017</th><th>2018</th><th>2019</th></tr>
<tr><td>Experiment 1</td><td colspan="5">Train set</td><td>Test set</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Experiment 2</td><td></td><td colspan="5">Train set</td><td>Test set</td><td></td><td></td><td></td></tr>
<tr><td>Experiment 3</td><td></td><td></td><td colspan="5">Train set</td><td>Test set</td><td></td><td></td></tr>
<tr><td>Experiment 4</td><td></td><td></td><td></td><td colspan="5">Train set</td><td>Test set</td><td></td></tr>
<tr><td>Experiment 5</td><td></td><td></td><td></td><td></td><td colspan="5">Train set</td><td>Test set</td></tr>
</table>

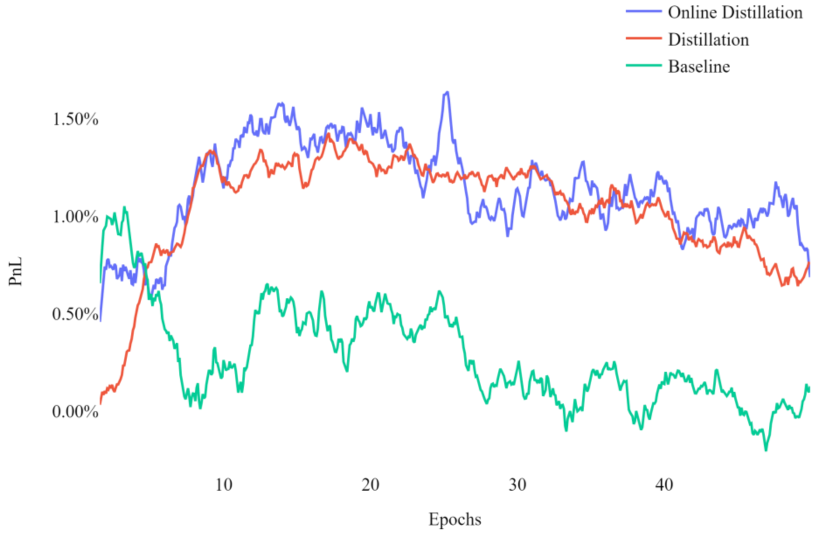
Out of sample predictions

The presented Results are an average of the test metrics of all 5 sub-experiments.

ARISTOTLE UNIVERSITY OF THESSALONIKI

Results Comparison





Conclusions

Conclusions

- Developed an end-to-end **Ensemble Based Online Knowledge Distillation** scheme that effectively mitigates the instability problem.

- A Teachers' ensemble can successfully be used to extract the new ground truth annotations.

- The proposed method can successfully be used to mitigate the intense instability and overfitting issues in financial trading.

Future research directions:

- Self-distillation approaches can potentially reduce the impact of noisy handcrafted labels on teacher models and further improve the performance of the method.

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 *<u>deepfinance.csd.auth.gr</u>*

Co-funded by Greece and the European Union

ARISTOTLE
UNIVERSITY OF
THESSALONIKI

Thank you for your attention!

Questions?

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

DeepFinance