

Sequence-to-Sequence Learning of Financial Time Series in Algorithmic Trading

Philip Arvidsson philip@philiparvidsson.com

Tobias Ånhed anhedtobias@gmail.com

December 2, 2016

Contents

1 Introduction	2
1.1 Background and motivation	2
1.2 Current research	2
1.3 Problem statement	3
1	

Chapter 1

Introduction

Time series forecasting—predicting future values of variables within a domain—is a largely unsolved problem in complex or chaotic domains such as weather (e.g. humidity, temperature or wind speed) and economics (e.g. currency exchange rates or stock prices). In this thesis, we will attempt to solve the problem by using LSTM-based (Long Short-Term Memory) RNNs (Recurrent Neural Network), something that has not been done before.

1.1 Background and motivation

Past attempts have approached the problem with algorithms that are not designed specifically to handle sequential datasets and instead assume that the data points are IID (Independent and Identically Distributed). Time series sequences have a temporal component dictating their intradependence and distribution within the dataset, where each element in the sequence is correlated to previous (with respect to time) elements through a plethora of factors.

The problem should be examined with algorithms, designed specifically for time series sequences, that take into account the temporal aspect of the datasets.

1.2 Current research

In the field of machine learning, HMMs (Hidden Markov Models), DBNs (Dynamic Bayesian Networks), TDNNs (Tapped-Delay Neural Networks) and RNNs (Recurrent

Neural Networks) are commonly used to handle sequential datasets.

RNNs (Recurrent Neural Networks) have historically had problems with vanishing (or exploding) gradients, but this problem was solved through the introduction of the LSTM-based RNNs (?). More recently, deep LSTM-based RNNs have successfully been applied to predict future time series sequences from historical, sequential datasets; such as predicting the weather for the next twenty-four hours (?), or predicting the next sequence of words in natural language (?).

1.3 Problem statement

Financial time series are used ubiquitously in algorithmic trading. In algorithmic trading, it is imperative that accurate predictions are made about numerous variables (such as volatility and returns) in order to time market entry and exit.

Since deep LSTM-based RNNs have only been applied within the algorithmic trading domain to a minimal extent, and since they have shown great success in solving similar problems in other domains, it raises the question whether the technique can be used to predict a future sequence of financial variables that can be used to time both entry and exit positions within a certain time horizon.

Presuming that correlations exist along the temporal dimension of the dataset, the problem is reduced to a matter of finding an appropriate set of features enhancing the correlations, on which to train the LSTM-based RNN.

Hur fungerar inledningens övergripande struktur? Finns det en röd tråd, och hur ser den ut?

Tycker att inledningen va bra på så sätt att det var en tydlig "cliffhanger" texten hänger ihop men känns väl som att det hade kunnat vara lite djupare förklaringar i bakgrunden kanske.

2. Hur fungerar de källor som åberopas. Är det en bra blandning, för mycket av den ena eller den andra, "tror man" på författarna, och är det någon röst man saknar som läsare?

Saknar överlag källhänvisningar i texten även om källhänvisningar finns så vet man inte som läsare vilka påståenden som är kopplade till vilken källa.

3. Behöver problemformuleringen avgränsas mer? Är de begrepp och teorier som är relevanta för problemformuleringen tydliggjorda i den tidigare texten.

Vissa algoritmer "namedroppas" i er text som jag antar är relevanta för er studie men de förklaras inte någe vidare så det är svårt för mig som läsare att förstå varför dessa är relevanta eller va det är som gör dom relevanta. Avgränsningen görs ganska tydligt implicit liksom i erat "problem statement", men kanske hade kunnat göras tydligare.

4. Vilken metod och vilken design verkar passa till den problemformulering som finns?

Design forskning

5. Hur fungerar texten som sådan. Vilken stil har författarna? Skriver de distanserat (använder inte vi), är språket (för) högtravande, (för) vardagligt. Vilka skrivråd vill ni ge författarna

Det är ganska distanserat jag skulle väl ge som råd att typ förklara mer varför exempelvis algoritmerna ni hänvisar till i bakgrunden är intressant relaterat till er studie, kanske annat än "are commonly used to handle sequential datasets" för jag antar att det också finns andra anledningar som gör dom relevanta till er studie?

Bibliography

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1733–1780, November 1997.

I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks, 2014.

M. A. Zaytar and C. E. Amrani. Sequence to sequence weather forecasting with long. *International Journal of Computer Applications*, 143(11), June 2016.