SysEng 5211/ElecEng 5310/Comp Eng 5310

Name: \_\_\_\_\_Ryan Patton\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Homework 2

This is an individual homework assignment.

The premise of the paper written by Kumar, Jain, and Singh was to survey how the stock market worked, and computational intelligence’s (CI) current uses in the field. The authors mentions six points of view for stock market forecasting based on computational intelligence methods: (1) the stock market analyzed and the related datasets, (2) the type of input variables investigated, (3) the pre-processing techniques used, (4) the feature selection techniques to choose effective variables, (5) the forecasting models to deal with the stock price forecasting problem, and (6) performance metrics utilized to evaluate the models. Building on this, the authors aimed to provide researchers/financial analysts with methods for approaching the development of intelligent methodologies applied to forecasting the stock market. The paper is broken down by the following sections: introduction, previous related work, terminology of stock market and computational intelligence, forecasting work flow (including stock market surveyed, input variables, data pre-processing, feature selection and extraction, forecasting models, hybrid prediction models, and performance evaluation), proposed work, conclusions, and references. The flow of the paper guides the reader through familiarization with computational intelligence and the stock market through applications, and finally builds on current applications to hint at what’s to come in CI’s applications in the stock market.

During Section 4’s discussion of the various forecasting work flows, the authors made sure to include a wide breadth of papers. The authors touched on a wide variety of topics which is difficult to do for such a survey paper on a broad topic with many niche CI applications. Although stock market forecasting primarily focuses on Artificial Neural Networks (ANNs), the authors made sure to give plenty of space to describe fuzzy logic and genetic algorithms (Gas). For a survey paper, I thought the paper fixated in on work people had done too much after describing each of the techniques for ANNs and fuzzy logic/GAs instead of bolstering a better transition. For example, the authors did a good job describing what all sorts of models were but they skipped over the theory/application behind all of them and some crucial math equations before jumping into past work to see how it is actually done. This part of the paper can be the most tiresome to work through but it also helps establish better credibility of their background research. Granted, it is a survey paper outlining such far-reaching interrelated topics making the nature of the paper difficult to tackle, but the paper could be considered lacking if a reader wanted to actually implement their own ANNs for example to maybe duplicate the work done.

The research methodology was a survey paper so it was meant to provide a wide variety of papers detailing the current state of affairs in the field and how the field has reached the point it has today. Their groupings of papers throughout seemed to be lazy to me as on page 9 they simply listed about 85 references after stating they were presenting a list of various stock market authors who have investigated and acquired data for evaluating the performance of their models along with the corresponding stock exchange. The authors do return to do some more in-text citations but the fact that they had originally listed out so many references in a row leaves me questioning whether they thought that was the most effective way to capture all the research they had done or if it was a simple way to cover their own research without focusing on that aspect of the paper. The formatting of their work almost invites investigation to see what content they had cited and its relevance. The majority of their research cited was for the terminology of stock markets/CI and following stock market models/research which makes sense. The number of references they had total, 162, should leave the reader with a good feeling that they had truly tried to capture a comprehensive survey approach backed by a lot of research and wide background of papers. The research papers cited were overall grouped by what CI method they were talking about, like GAs, hybrid GAs, ANNs, etc… which seemed to logically flow well.

The results touched on what the authors considered performance measures for various models that best captured stock market forecasting. They start out by stating the obvious, that the performance and robustness of the trained models depends upon the type of problem being modeled and how it could be classified. They touch on variables affecting the type of problem like bullish/bearish trends, precision, recall, F-Score, confusion matrices, joint prediction error (JPE), stock index, price predictions, closeness between actual and predicted values, average relative variance, etc… The paper was certainly not lacking in the amount of accounted-for variables depending on the problem. From their various studies, the results support hybrid models predicting more accurately than traditional approaches grouped off into one category. For example, a model using ANNs but applying evolutionary optimization techniques to tune parameters performs better than simply using an ANN framework. The authors felt confident in using technical indicators to construct feature vectors for stock market prediction. Their studies flow well from the introduction of the paper because it touches on how important technical indicators can be for a model’s success and the pros/cons of each type of CI method used. Their hybrid approach only makes sense and obviously would be heavily dependent on the type of problem being investigated. Even though this paper was published in February 2020, their results do leave me with two looming questions: with the constant evolution of CI, how do they expect models to change in the coming years and how do they anticipate COVID-19 affecting the US stock market long-term?

The authors concluded that they achieved what they set out to do, “survey the important and up-to-date contributions in the domain of computational intelligence to solve stock market forecasting problems.” The authors and readers learned a lot from their technical analysis on pre-processing, dimensionality reduction, and forecasting future trends or predicting future stock prices. They confirmed what they anticipated before conducting the survey that CI can be effectively utilized in stock market forecasting and that systematic approaches for the development of intelligent methodologies can lead to better forecasting. Two other intended goals of the authors were to discuss basic stock market terminologies and CI approaches and to present sources of data to obtain historical stock market data of both national/international markets, both of which the authors went above and beyond doing. The authors assert that they are the first survey to present these sources of data in the stock market domain. Finally, the authors wrap up the survey with four major findings: (1) technical indicators play a prominent role in stock market forecasting, (2) identifying pre-processing and feature selection techniques help improve accuracy, (3) CI approaches can be effectively used to solve stock market forecasting problems with high accuracy with hybrids being the most effective, and (4) performance metrics are not unique, supporting the use of different combinations of metrics by different authors from their data sources.

Overall, I thought their research was well-founded and provided plentiful data to back up their conclusions. I think it can be hard to draw conclusions when using ANNs because it is sort of a black box but their assertion that hybrid models are most effective makes sense. Their investigation of stock markets in the United States and other countries is interesting to me, as many papers seem to only focus on the United States, which does not necessarily represent how other countries’ stock markets perform. I was very impressed with the way the authors broke down all of the information and thought the information flowed very logically. They did a good job hitting on key terminologies/topics without making the paper too low-level and technically intertwined to escape the point of a survey paper. When CI is brought up, especially when talking about the stock market, I think it can be easy to glance over EAs and fuzzy logic and jump straight to giving ANNs more attention so it was refreshing to see how they too played a role. I believe the paper was written by non-native English speakers so the grammar is poor at times but I was still able to derive the meaning from any sentence. Some charts/graphs/tables seemed to be thrown in as space-fillers without strategically contributing to the points they were trying to make and were very easy to glance over without gaining a better understanding. The authors devoted 1-2 pages on data pre-processing but I do not think this was sufficient due to how much data pre-processing affects the health/performance of models. I understand it is hard to capture pre-processing when they then analyze different peoples’ studies where they have to revert to talking about how they applied pre-processing but it is hard to grasp the importance of pre-processing from their paper as they convey more that it is just another step in using ANNs. Overall, I was very pleased reading their paper and came up with many key takeaways from it as well, as I intend to jump into my own survey paper on autonomous driving.