INTRO PART

Previous research such as (Andridge & Little, 2011) [1] analyzed the use of Hot-deck imputation where the focus was on finding the best way to apply the hot deck and obtain inferences from the completed data set. (Che, 2018) [2] and (Cao et al., 2018) [3] focused on the performance of Recurrent Neural Network where the goal was to capture the long-term temporal dependencies in time series, but also utilizes the missing patterns to eventually achieve better prediction results. Several methods have been tested for time series data imputation. For example (Moritz et al., 2015)[4] compared linear interpolation and mode, median, mean and concluded that linear interpolation was the best suited for imputing data. These previous researches have been used as a guidance but not directly further improved on. Where hot-deck was analyzed for the best way to be applied from a completed data set, this research focused on which type of gap size hot deck should be used for.

METHODS PART

Hot deck is a well received imputation method under practitioners and has a reason for its popularity. It mainly avoids the issue of cross-user inconsistency that can occur when analysts use their own missing-data adjustments. Apart from this, hot-deck doesn’t rely on model fitting for a variable to be imputed. This makes it less sensitive to model misspecification. Lastly there is a reduction in non-response bias. A disadvantage however is that the hot deck makes implicit assumptions to match donors to recipients, so it is everything but assumption free (Andridge & Little, 2011).

Recurrent Neural Networks have strong properties such as strong prediction performance as well as being able to capture long-term dependencies and variable length observations. The main advantages of RNN are the fact that it is dynamic, meaning it can be used in many temporal processing models and applications. Rnn’s also possess universal approximation property, this means that they are capable of approximating arbitrary nonlinear dynamical systems with arbitrary precision, by realizing complex mappings from input sequences to output sequences. (Sharkawy, 2020) [4] However, there have not been works which design RNN structures incorporating the patterns of missingness for time series classification problems. (Che, 2018) [2]

[1]

Andridge, R. R., & Little, R. J. A. (2011). A Review of Hot Deck Imputation for Survey Non-response. *A Review of Hot Deck Imputation for Survey Non-response*.

[2]

Che, Z. (2018, 17 april). *Recurrent Neural Networks for Multivariate Time Series with Missing Values*. Nature. Geraadpleegd op 20 december 2021, van https://www.nature.com/articles/s41598-018-24271-9

[3]

Cao, W., Wang, D., Li, J., Zhou, H., Li, L., & Li, Y. (2018). BRITS: Bidirectional Recurrent Imputation for Time Series. *BRITS: Bidirectional Recurrent Imputation for Time Series*.

[4]

Sharkawy, A. N. (2020). Principle of Neural Network and Its Main Types: Review. *Journal of Advances in Applied & Computational Mathematics*, *7*(1), 8–19. https://doi.org/10.15377/2409-5761.2020.07.2