

Analysis of Cryptocurrency Commodities with Motifs and LSTM

Benjamin Barry and Martin Crane

School of Computing, Dublin City University, Dublin 9
benjamin.barry5@mail.dcu.ie, martin.crane@dcu.ie

Abstract. Time Series Analysis has long been used as a method to help predict future events based on previous data. This area is well-defined and studied but there are still areas of Time Series Analysis that are problematic. One such problem is isolating patterns in highly volatile datasets over long time intervals. In an effort to help with such issues e.g. those seen in fluctuating financial datasets, a new approach is needed. Borrowing its name from Music and more recently Computational Biology, Motifs are small repeating patterns in datasets. Using these subsequences, further forecasting can be improved as Motifs can be shown to aid algorithms which require certain initial conditions. This method will potentially lead to new insights in the study of cryptocurrencies. This paper will present a case where an optimized Motif length will be used to aid the prediction of Bitcoin through the use of an LSTM Neural Network, yielding an 8% decrease in RSME for one test case.

Keywords: Time Series Analysis · Cryptocurrency · Motifs · LSTM Neural Networks

1 Introduction

Forecasting has derived many methods for calculating the future behaviour of a system. Statistical methods, for instance, look to predict the relationship between variables when they are reduced to a simplistic expression. The area of statistical forecasting looking at the effect time has on a system is known as *Time Series Analysis* and this area will form the core of this work. Most aspects of forecasting Time Series data revolve around finding patterns within said data, as an attempt to see if these patterns can be projected into the future. As put forward by [12] Time Series problems can be reduced down to $\text{data} = \text{pattern} + \text{error}$. Therefore, if we can isolate the patterns and reduce the error (or noise) in a dataset we can decompose a series into something we can extend forward in time. However, what happens if the data has little to no “obvious” patterns? Here traditional methods can begin to fail and the picture they project becomes less accurate and hence less reliable. A new method is needed to try and better model these volatile systems, a change in point of view from the big to the little. Using improvements in computational power, we are no longer limited to looking at coarse granularity to try to model patterns. We can now begin to look at microscopic intervals and

a new world of pattern repetition becomes apparent. In this paper, we look at bringing together the power of time series Motifs to optimize Neural Network’s (NNs) ability to recognise patterns.

In Section 2 we examine previous research in the areas of Motifs with a special focus on the area of matrix profiles. We also examine prior work in Bitcoin forecasting as well as looking at some inherent challenges presented. In Section 4 we detail the results of the work in terms of Similarity Scores and Occurrences before looking at the overall accuracy and a comparison between LSTMs with and without Motifs. Section 4.4 discusses these results, while Section 5 concludes with some possible future directions.

2 Related Work

2.1 Motifs

Certain bodies of work have introduced Motifs (unknowingly) as opposed to in a theoretical framework for their study and evolution. For this, we need to look to three particular papers that represent a linear progression of the work.

Although there have been several co-writers across the three papers [11], [2] and [14], Eamonn Keogh features heavily in all and he can be considered the instigator of the field. The fundamental concept of Motifs is encapsulated by Keogh in [11] where he defines a Motif as the subsequence within a time series which is most similar to the most other subsequences in that series, $D(C_i, C_j) > 2R$, for all $1 < i \leq K$. Here R is a numeric value picked using a Distance Measure such as the Euclidean norm. A consequence of the definition is that all C_i are mutually exclusive. In Figure 1, A, B and C are a single repeating structure found

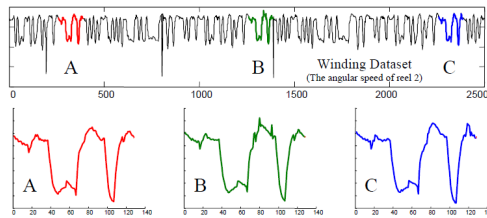


Fig. 1. A Motif in Time Series data from [2].

at different points in the time series. This can be seen to be nearly identical in structure by eye and when compared mathematically they are seen to fall within our distance range to be called the “same” i.e. our R from above.

2.2 Matrix Profile Work

The goal of most research within the field of Matrix Profiles, is to reduce the run time of the Brute Force approach ($O(n^2m)$ [11]). The method known as

STAMP (*Scalable Time series Any-time Matrix Profile*) uses the Fast Fourier Transform to reduce the computation time to $O(n^2 \log n)$ [2]. The method used here is STOMP (*Scalable Time series Ordered-search Matrix Profile*) which runs in $O(n^2)$ [10]. In this algorithm, through algebraic changes that better use the component order and invariance of the matrix, run time is improved.

The algorithm uses Piecewise Aggregate Approximation to reduce dimensionality and then Discretization is performed to transform the time series into “Words” which can be easily compared. This allows for a more efficient calculation of distances exactly [14] (unlike other methods).

2.3 Bitcoin and its Modelling Challenges

Operating under the pseudonym “Satoshi Nakamoto”, a developer published a paper [15] where they propose the idea of a Cryptocurrency. From that point onwards the rise of cryptocurrencies has been exponential and highly volatile ¹. The rise and fall of Bitcoin has led many to try to predict what it will do next [4]. The price of Bitcoin has been modelled by, among others, Krystoufek [9] who found by studying Bitcoin Price according to fundamental Economic laws, the price is not ever-increasing. Since 2018 Bitcoin is close to its model-implied price. This finding would seem to back up the notion that Bitcoin had entered a *Mature Phase* [4]. Due to the (relatively) low amount of trading done on Bitcoin pre-April 2017, the price remained less volatile. With Bitcoins increased popularity, its value changed wildly from day to day. Hence Bitcoin is best considered in 2 ages, its Immature state and its Mature state. As the price of Bitcoin in the Immature state has been found to be almost perfectly correlated with other Cryptocurrency prices [3], this has given rise to arbitrage opportunities which have been pursued as a means to find the “next” Bitcoin. Further investigation of the variability of Bitcoin price has taken the form of research into existence of regime changes in the price, [8] found that Recurrent Neural Networks outperform Hidden Markov Models at predicting volatility spikes.

Other authors have also found challenges due to the extreme fluctuations seen in the Bitcoin market over small increments of time. For example in July 2017, its value increased 25% in a single day. Previous research has attempted to solve the problem using the correlation between the number of transactions and price. This yielded positive results for Bitcoin returns without, however, being able to predict volatility [1]. The issue with this approach is the need to know the number of transactions in order to create a prediction. A method is needed that can model forward the market with less restrictive input parameters.

A statistical approach is taken by Shah and Zhang [16] where they use Bayesian Regression to model future values of Bitcoin. They achieve good results with this approach and are able to nearly double their initial investment in less than a 60-day period.

LSTM RNNs have been found to be an efficient way to predict Bitcoin price from historical prices and moving average technical indicators. Other Machine

¹ Its max value over \$18,000 to current \approx \$8,000 (coinbase.com/price/bitcoin).

Learning methods, e.g. [13] found LSTM RNNs produced the highest accuracy on price prediction. [6] used LSTM model to predict Bitcoin price with, as model inputs, macroeconomics factors such as global currency ratios and blockchain information. [7] have used them to mine news articles and tweets to infer the relationship between their information, certain commodity prices and Bitcoin price direction yielding an accurate prediction technique.

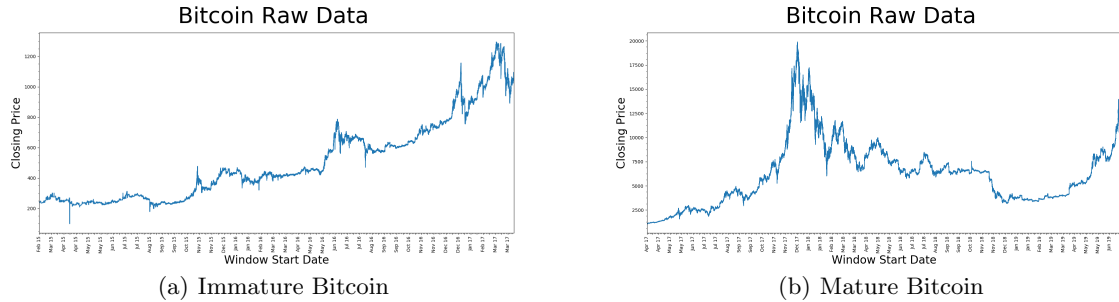


Fig. 2. Historical Bitcoin Graphs

3 Methods

3.1 Matrix Profiles

What underpins the Keogh algorithms [11], [2] and [14] is the concept of a *Matrix Profile*. In essence, this reduces a time series T into a matrix where each cell represents the distance between a subsequence and all other subsequences of equal length in the time series. This produces a matrix of size N by N (N is the length of the time series T less the length on the subsequence). For a time series T with 100 data points the Matrix Profile for all subsequences of length 10, would consist of a 90 by 90 matrix. A common metric used is the z-normalized Euclidean distance due to its ability to account for variance and mean. Consider Figure 3, where the blue and red lines represent 2 different time

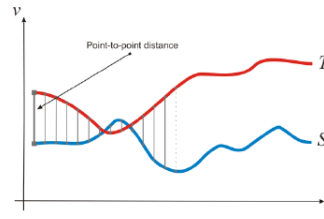


Fig. 3. Euclidean Distance between subsequences

series T and S . Here the 2 series are mapped on top of each other over a period

of time with arbitrary values on the y -Axis. The vertical lines show the distance between corresponding points on each series (i.e at the same point in time). The modified sum of these distances represents the Euclidean distance at a simplified level. For Motifs, we are only interested in the lowest distance value i.e. the most similar subsequence.

3.2 Data Used

Three separate time granularities were chosen for both the Mature and Immature data within Bitcoin. Since the Poloniex² data source was accurate to 5-minutes, this was the lowest granularity picked. From here, 15-minute and 30-minute intervals were also chosen. This data counts are shown in Table 1. Graphs of the Mature and Immature Bitcoin data are shown in Figure 2.

Table 1. Dataset Size

Dataset	Granularity	Size
Immature	5 Minutes	221,829
	15 Minutes	73,937
	30 Minutes	36,965
Mature	5 Minutes	236,421
	15 Minutes	78,789
	30 Minutes	39,381

4 Results

4.1 Matrix Profile Outputs

A sample of the standard STUMPY algorithm outputs are seen in Table 2. This table shows the Starting position of a subsequence (*Index*), the Euclidean Distance (*Similarity*) to its closest copy, the starting index of where that copy is (*Start*) and how often the pattern occurs (*Occurrence*).

Table 2. 3 Matrix Profile - Immature 5 Minutes

Index	Similarity	Start	Occurrence
0	0.33437	14292	27
1	0.07794	3319	472
2	0.07794	4329	398
3	0.07794	9592	221

² <https://poloniex.com/>

4.2 Matrix Profile Graphs

The data in Table 2 can be used to create an overall result for all granularities within the data set graphically. Figure 4 shows the Similarity (Euclidean Distance) of each subsequence of length 3, 6, 12, 24 and 48. With Figure 4(e) as one of the clearest examples, we can see peaks within the graph or “Discords”. These can be considered the opposite of Motifs i.e. they represent seldom repeating occurrences or anomalies. The lower the value on the y -Axis the closer the Motifs are to each other, i.e. over larger windows we can see the Similarity Score increasing so the Motifs are becoming less similar.

4.3 Long Short Term Memory (LSTM)

Due to the large number of points in the input data (Immature 15-minutes and Mature 15-minutes) and limited hardware resources³ the LSTM NN takes a lot of time to run. The computation times were noted as 23.1 hours, 7.5 hours and 26.6 hours, 8.4 hours on the Mature and Immature dataset running the Motif and the standard algorithm respectively. A comparison can be made between our version (with Motifs) and a standard 1 point LSTM NN in Figures 7 (a) & 8 (a) for the Mature and Immature cases respectively. Table 3 show Mean Square Error for the Trained and Tested case for Mature and Immature for Motifs with LSTM NN and the 1 point LSTM NN.

Table 3. LSTM Mean Square Error

Dataset	Type	Train/Test	MSE
Immature	Motif	Train	92.35
		Test	247.78
	Standard	Train	136.22
		Test	501.27
Mature	Motif	Train	2,095.82
		Test	1,689.28
	Standard	Train	2,761.33
		Test	1,832.49

4.4 Discussion

Optimum Motif Length The goal from analysis of the outputs is to attempt to find an “Optimized” Motif that we can then pass into the LSTM to try to improve its accuracy. To do this, it was best to compare the Motif lengths across

³ The code was run on an Intel i7-5600 CPU @ 2.60GHz with 8GB of RAM

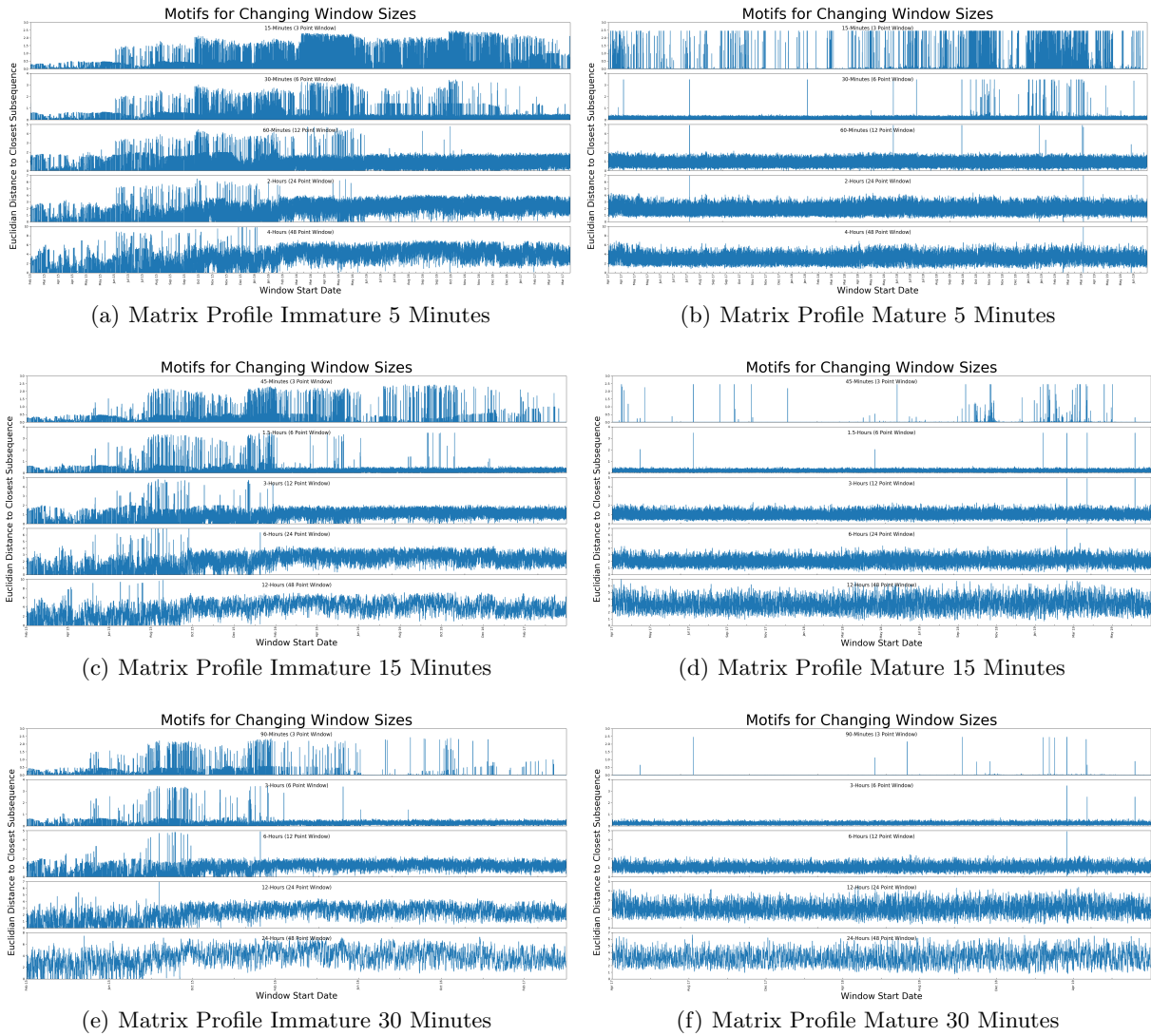


Fig. 4. Matrix Profile for Different Granularities

two metrics: Similarity Scores and Motif Occurrence. Hence, to establish which Motif length to input, each will have to be taken into account and an overall best value will need to be selected. The answer from this section mixes the qualitative and quantitative. We need to visually interpret graphical outputs, in conjunction with the underlying numeric scores. As such, the optimum Motif length is open to interpretation.

Similarity Score For both the Mature and Immature dataset for the five subsequence lengths tested, an output as seen in Table 2 was created. The longer Motifs tend to get less similar (this makes sense intuitively with more points to calculate the Euclidean distance between). The pitfall at this point would be to just consider the length 3 Motif as the best option due to it having the smallest mean Euclidean distance. However, we must consider the reason behind this. Figure 5 contains the graphed Similarity Score (in intervals of 0.5 where the final value on the x -Axis contains all values above it). All these graphs show the dominance of the zero score Motifs. These occur often at low granularity due to minute differences in Bitcoin’s value. These effectively are straight lines in the graph. This behaviour is dominant in the 5-minute 3 and 6 length Motifs (it is also dominant in most of the Immature data due to the lack of trading at the beginning of Bitcoin’s life). We need to find the best compromise between high Similarity Scores and the dominance of the straight-line “0” score. In all six graphs in Figure 5, the blue (3 Motif) and the orange (6 Motif) plots are dominated by the 0 score Motifs, while the purple (48 Motif) plot occurs more often at the higher values. The green (12 Motif) plot and the red (24 Motif) plot are our best compromise. Of note in these graphs is the apparent Gaussian distribution sitting on the spread of some scores (this is most obvious for the red (24 Motif) plots). This may be explained using the Law of Large Numbers [5]. Due to the lack of dominance of the 0 score and as there are fewer larger scores, the values are tending towards their Expected Value. This effect is more pronounced in the Mature dataset due to the movement of Bitcoin towards a traditional commodity [4]. To split our choice between a 12 Motif and 24 Motif we consider the Occurrence counts of our Motifs.

Motif Occurrence When we speak of the “Occurrence” of a Motif we are referring to how often it appears in a dataset. If we look back to Table 2 we can explain the field as how many times the Start, i.e. the starting index of the repeating subsequence relative to the Index, occurs. This is to say that the subsequence at Index 0, along with 26 other subsequences are most similar to the subsequence at Index 14292. This means the Motif repeats throughout the time series multiple times. Although all 27 subsequences may not have the same Similarity Score as 0, they will form a Motif. Highly repeating structures are again usually akin to straight lines and often, after a further look, are uninformative. Due to the high number of possible values for the Occurrence in the graphs, the values need to be bucketed: Anything with an Occurrence value of 1 or 2 is kept as is, whilst 3-5, 5-10 and 10-50 are grouped and finally anything above 50 are

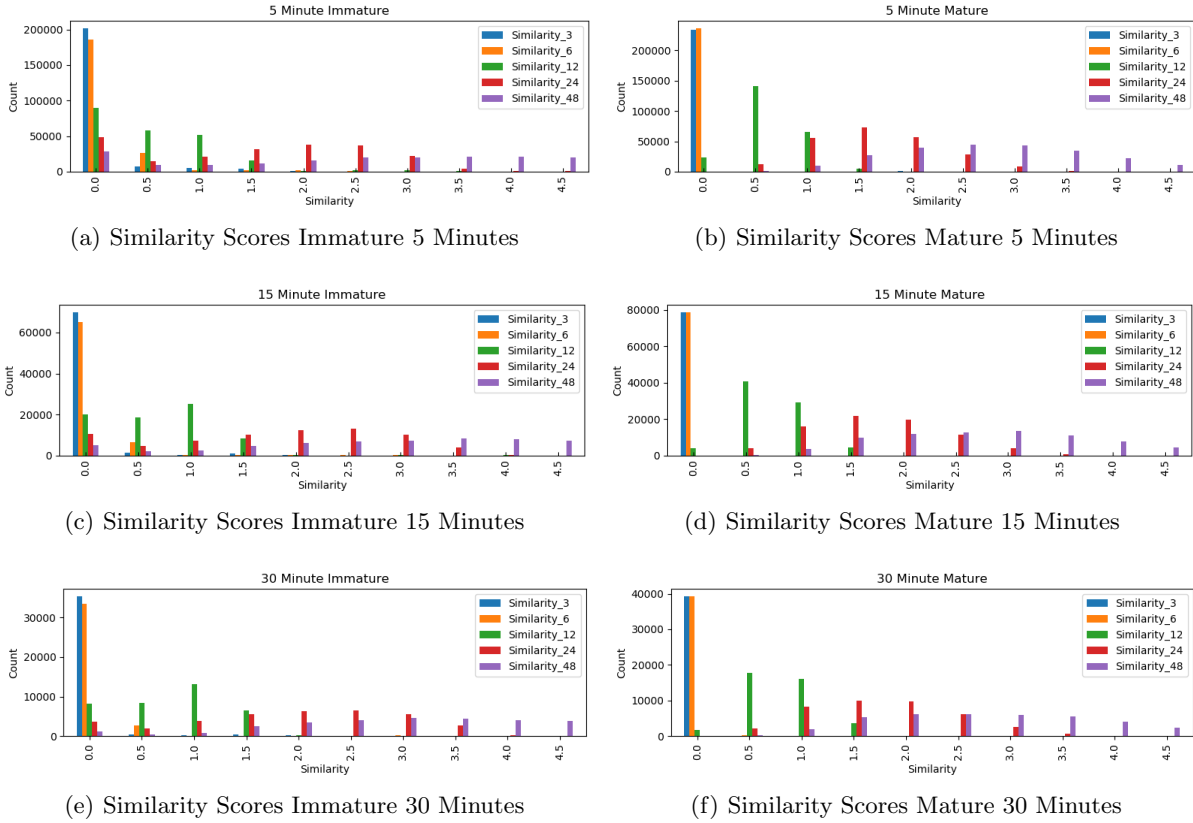


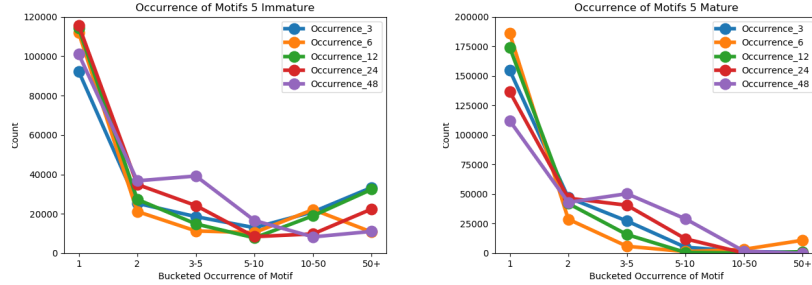
Fig. 5. Similarity Scores for Different Granularities

included in the 50+ bucket. Subsequences with an Occurrence value in these categories are counted and used as the y -Axis value. In Figure 6, we are interested in the green (12 Motif) and red (24 Motif) plots from analysis in Section 4.4. From the graphs there is little to choose between the two Motif lengths. In general, the green plot seems to have more of the undesired Occurrence values of 1 and Occurrence values in the 50+ range. Hence, we can select the 24 Motif as the best option between the two and proceed to using it in the LSTM NN.

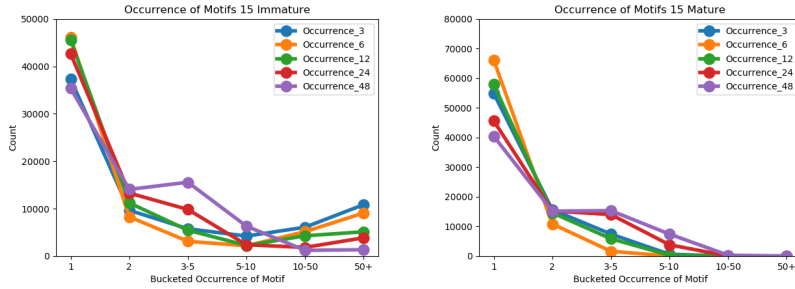
4.5 LSTM Accuracy

The LSTM NN was not optimized as that was not the goal of this paper, as such, some of the hyperparameters chosen are not useful in a real world scenario. This is especially true of the batch size of 1. An increase here would have reduced the computation time. In this sense there is little to no value in using the computation times as a metric to measure if the Motif using LSTM outperforms a traditional LSTM.

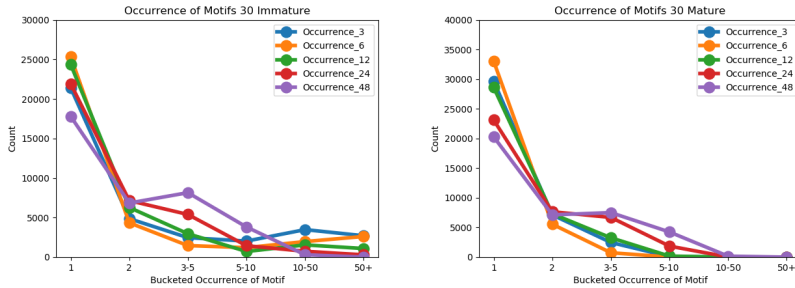
The simple goal of the LSTM NN is to see does using Motifs as an intermediary step between raw data and model help our outcome. Thus Figures 7 (a) and (b) let us compare our Motif version to the standard version of the algorithm for the Mature dataset. It is clear that the former tracks its training data much more effectively. What is very obvious is that after only 25 epochs the test outputs



(a) Motif Repetition Immature 5 Min- (b) Motif Repetition Mature 5 Minutes



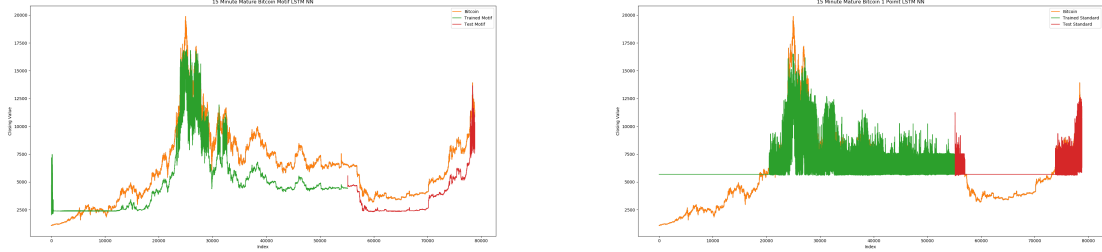
(c) Motif Repetition Immature 15 Min- (d) Motif Repetition Mature 15 Minutes



(e) Motif Repetition Immature 30 Min- (f) Motif Repetition Mature 30 Minutes

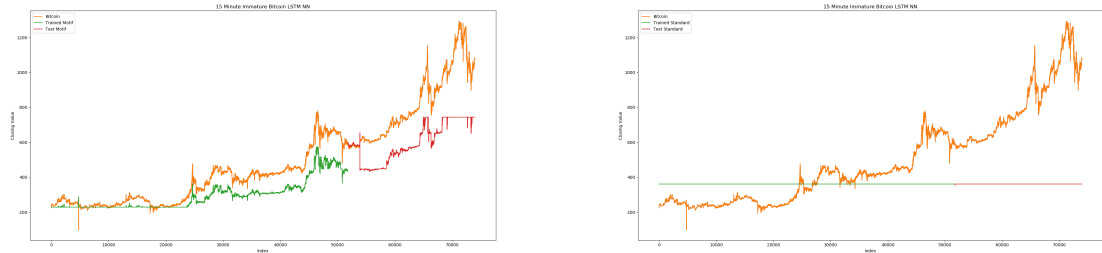
Fig. 6. Motif Occurrence for Different Granularities

trend very closely to the true values, just shifted down in value by roughly 750. Table 3 shows a reduction in the test RMSE by 8% and training of 24%. There are significant computation time increases (66%), but with script optimization and better hardware this could be reduced. Figures 8 (a) and (b) show the effect introducing Motifs has to the Immature dataset. Table 3 shows us that there is a reduction in the test RSME by 33% and training of 50%.



(a) LSTM NN for Mature Bitcoin using Motifs (b) LSTM NN for Mature Bitcoin not using Motifs

Fig. 7. LSTM NN for Mature Bitcoin With and Without Motifs



(a) LSTM NN for Immature Bitcoin using Motifs (b) LSTM NN for Immature Bitcoin not using Motifs

Fig. 8. LSTM NN for Immature Bitcoin With and Without Motifs

5 Conclusion

The goal of this work was to study the application of Motifs in pattern detection in volatile data. The computation speed and exactness of the Matrix Profile make it a powerful tool and results back this. Future work could be to further integrate Motifs into LSTMs by putting Motifs in the individual NN nodes. If, training a LSTM, a recognised Motif occurred, forcing the node to abandon its standard approach of randomizing the weights matrix and simply complete the Motif could help our learning rate. This would require a more proficient NN to be built to better handle the computations and reduce run-times.

There is the potential that capturing enough Motifs in a sequence could replace

a series with a Motif piecewise reduction, reducing the total information stored by potentially multiples of bits. All these are possible future directions.

The work presented here is a simple use of Motifs in a practical way. The results indicate that there is strong potential to embed Motifs in different areas of Computer Science.

6 Acknowledgements

The authors are grateful to Prof. H. J. Ruskin for her helpful comments. This research was (partially) conducted with the financial support of SFI under Grant Agreement No. 17/SP/5447 at the ADAPT SFI Research Centre at DCU. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant no. 13/RC/2106.

References

1. Balcilar, M., Bouri, E., Gupta, R., Roubaud, D.: Can volume predict Bitcoin returns and volatility? *Economic Modelling* **64**, 74–81 (2017)
2. Chiu, B., Keogh, E., Lonardi, S.: Probabilistic discovery of time series motifs. In: Proc. of 9th Int. Conf. on KDD, Washington, DC, Aug 24 - 27 (2003)
3. Cretarola, A., Figà-Talamanca, G.: Detecting bubbles in Bitcoin price dynamics via market exuberance. *Annals of Operations Research* (Jul 2019)
4. Drozd, S., Gebarowski, R., Minati, L., Oswiecimka, P., Watorek, M.: Bitcoin market route to maturity? *CoRR* **abs/1804.05916** (2018)
5. Feller, W.: The strong law of large numbers. In: *An introduction to probability theory and its applications*, vol. 1(3), pp. 243–245. Wiley (1968)
6. Jang, H., Ko, H., Lee, J., Lee, W.: Predicting bitcoin prices by using rolling window lstm model. In: Proc of KDD Data Science In Fintech Workshop (DSF) (July 2018)
7. Kinderis, M., Bezbradica, M., Crane, M.: Bitcoin currency fluctuation. In: 3rd Int Conf on Complexity, Future Information Systems and Risk, 20-21 Mar (2018)
8. Kodama, O., Pichl, L., Kaizoji, T.: Regime change and trend prediction for Bitcoin time series data. In: CBU Int. Conf. Proc. vol. 5, pp. 384–388 (2017)
9. Kristoufek, L.: Is the Bitcoin price dynamics economically reasonable? *Physica A* p. 120873 (2019)
10. Law, S.M.: STUMPY: A powerful and scalable python library for time series data mining. *The Journal* **4** (2019)
11. Lin, J., Keogh, E., Lonardi, S., Patel, P.: Finding motifs in time series. *Proceedings of the Second Workshop on Temporal Data Mining* pp. 53–68 (10 2002)
12. Makridakis, S., Wheelwright, S., Hyndman, R.: *Forecasting methods and applications*. John Wiley & Sons (2008)
13. McNally, S., Roche, J., Caton, S.: Predicting the price of bitcoin using machine learning. In: 26th Euromicro Int Conf on Parallel, Distributed and Network-based Processing (March 2018)
14. Mueen, A., Keogh, E., Zhu, E., Cash, S., Westover, B.: Exact discovery of time series motifs. In: Proc 2009 SIAM Int conf on data mining. pp. 473–484 (04 2009)
15. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2008)
16. Shah, D., Zhang, K.: Bayesian regression and Bitcoin. In: 52nd Allerton Conf on Communication, Control, and Computing, Monticello, IL, Sept 30 - Oct 3 (2014)