Modeling symbolic candle chart time series

Heewon Park; Fumitake Sakaori

Abstract — This study introduces a new type of symbolic data namely a candle chart valued time series, and presents a new approaches of candle chart valued time series for forecasting of stock index direction (i.e. up and down) based on future candle chart form. From the approaches for interval valued time series, we propose forecasting methods for the candle chart valued time series based on a combination of two mid-point and two half range between the highest index and the lowest index, and between open index and close index. Also we propose new sum of squares for candle chart valued time series. To evaluate proposed methods, we describe forecasting result of real data set consisting of Asian major 5 countries’ stock market indexes. The forecasting results show that the new approaches and sum of square which are based on approach of interval valued time series outperform than others in forecasting candle chart.

Keyword: Candle chart; Symbolic data; Time series; interval valued data.

1 Introduction

With the development of computer and the data collection technology, size of database which should be dealt with is growing. For this reason, summary of information and visualization from the huge database are increasingly important. Along with this situation, the symbolic data analysis (Diday and Nori, 2008) has been introduced as extension of classical data analysis method to new type of data namely a symbolic data (e.g., interval valued data, histogram data, multimodal data, etc). The symbolic data analysis takes the information which can not be represented within the classical data model into account, and enable to summarize and visualize the huge data effectively.

In this study, we introduce new type of symbolic data namely a candle chart valued time series (CTS). The candle chart which is usually used for forecasting the direction of future stock index is composed with an open, close, highest, lowest stock index (or price). It means that we can consider CTS as a symbolic data since CTS is type of aggregated data. Also we propose forecast methods for candle chart which is extension of the approaches for symbolic interval valued time series. Using the forecasted candle chart, we can forecast direction of stock index (i.e, up or down).

2 Symbolic candle chart data

The candle chart has been widely used for forecasting direction of stock index. We can consider the candle chart which composed by four indexes, that is, open ($y_{t}^{o}$), close ($y_{t}^{c}$), highest ($y_{t}^{h}$) and lowest ($y_{t}^{l}$) index as a symbolic data.

A forecast method for candle chart time series can be considered using each 4 indexes, open, close, highest and lowest index. It is well known that forecasting methods using midpoint and half range superior to forecast accuracy in case of interval valued time series (Lima Neto and De Carvalho, 2008). Therefore, we may consider to use midpoints and half ranges between open and close index ($y_{t}^{oc}$, $y_{t}^{oc}$), and between the highest and the lowest index ($y_{t}^{hm}$, $y_{t}^{hl}$):

\[
\begin{align*}
    y_{t}^{oc} &= \frac{y_{t}^{o} + y_{t}^{c}}{2}, & y_{t}^{oc} &= \frac{y_{t}^{c} - y_{t}^{o}}{2}, \\
    y_{t}^{hm} &= \frac{y_{t}^{h} + y_{t}^{l}}{2}, & y_{t}^{hl} &= \frac{y_{t}^{h} - y_{t}^{l}}{2}.
\end{align*}
\]

*Department of Mathematics, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan, E-mail: heewon@math.chuo-u.ac.jp, Tel: +81-80-3936-1446

†Department of Mathematics, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
3 Modeling for forecasting candle chart valued time series

It is well known that the financial time series has the volatility clustering. The numerous studies of financial time series with volatility clustering have been progressed by the ARCH model (Chen and Jayaprakash, 2005). Therefore we consider the Autoregressive moving average (ARMA) with Autoregressive conditional heteroscedasticity (ARCH) model for the CTS. The ARMA(\(p, q\))=ARCH(\(s\)) model is given by

\[
\phi(B)y_t = \theta(B)\varepsilon_t + \eta_t, \\
\eta_t = \sigma_t \varepsilon_t,
\]

where the \(\eta_t\) are i.i.d. random variables with mean 0 and variance 1, which is independent of past realizations \(\eta_{t-1}\), and

\[
\sigma_t = \sigma_0 + \sum_{i=1}^{s} \alpha_i \eta_{t-i}^2.
\]

3.1 The 4-indexes method

We first introduce a 4-indexes method consist of the 4 indexes, the open (\(y^o\)), close (\(y^c\)), highest (\(y^h\)) and lowest (\(y^l\)) index time series. In the 4-indexes method, we apply time series model to the open (\(y^o\)), close (\(y^c\)), highest (\(y^h\)) and lowest (\(y^l\)) respectively. The sum of square for the 4-indexes method is given by

\[
S_1 = \sum_{i=1}^{n} (e_i^o)^2 + \sum_{i=1}^{n} (e_i^c)^2 + \sum_{i=1}^{n} (e_i^h)^2 + \sum_{i=1}^{n} (e_i^l)^2.
\]

3.2 The MMRR method

In this section, we proposed a Centre-Centre and Range-Range (MMRR) method similar to CRM method for interval valued time series (Lima Neto and De Carvalho, 2008). The MMRR method is composed with four time series as two midpoints (\(y_{oc}^m, y_{hl}^m\)) and two ranges (\(y_{oc}^r, y_{hl}^r\)) of interval between the open (\(y^o\)) and close (\(y^c\)) index, and between the highest (\(y^h\)) and lowest (\(y^l\)) index respectively, in the view point of symbolic data. We assume that \(y_{oc}^m \geq 0, y_{hl}^m \geq 0, y_{oc}^r \leq \infty < y_{hl}^r < -\infty\) and \(y_{oc}^l \leq y_{hl}^l \leq y_{oc}^h\). We also propose the three types of sum of squares for the MMRR method as in Lima et al. (2008) :

\[
S_2 = \sum_{i=1}^{n} (e_i^{ocm})^2 + \sum_{i=1}^{n} (e_i^{ocr})^2 + \sum_{i=1}^{n} (e_i^{hlm})^2 + \sum_{i=1}^{n} (e_i^{hlr})^2
\]

\[
S_3 = \sum_{i=1}^{n} (e_i^{ocm} + e_i^{ocr})^2 + \sum_{i=1}^{n} (e_i^{hlm} + e_i^{hlr})^2
\]

\[
S_4 = \sum_{i=1}^{n} (e_i^{ocm} + e_i^{ocr} + e_i^{hlm} + e_i^{hlr})^2
\]

\(S_3\) reflects the correlation between the midpoint and the half range of open and close index, and between the highest and lowest index. Also \(S_4\) reflects the correlation of two midpoints and two half ranges.

Also, we may consider extensions of hybrid models given in Maiaset al. (2008) to the symbolic candle chart data.

4 Applications: Asian major 5 countries' stock market index

As we mentioned above, the goal of this study is forecasting candle chart form as shown Figure1 and forecasting direction.

To compare the practicality of the proposed methods, we describe the Asian major 5 countries' (China, Hongkong, Japan, Korea, Singapore) stock market index based on the CTS. We forecast direction of index by
Figure 1: Candle chart and forecasted candle chart of Kospi 200

Table 1 shows proportions that forecast results of future stock index direction are right. We can see that the MMRR method with $S_2$ is superior to others. Table 2 shows root mean square-error and square of correlation coefficients. We can see that the MMRR method with $S_3$ outperforms than others.

From above results, we identify that the MMRR method outperform than the 4-price method in forecast accuracy.

5 Concluding remarks

In this paper, we introduce the new type of the symbolic data namely the candle chart valued time series composed with the open, close, highest and lowest index. Moreover, we proposed new methods and new sum of squares for fitting the time series model to the candle chart valued time series. From fitted valued of indexes included

<table>
<thead>
<tr>
<th></th>
<th>Kospi 200</th>
<th>Nikkei 200</th>
<th>Shanghai</th>
<th>STI</th>
<th>hansen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>50.1</td>
<td>50.0</td>
<td>46.4</td>
<td>57.9</td>
<td>31.6</td>
</tr>
<tr>
<td>Down</td>
<td>50.0</td>
<td>50.1</td>
<td>46.4</td>
<td>57.9</td>
<td>31.6</td>
</tr>
<tr>
<td>4-price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>44.4</td>
<td>51.9</td>
<td>64.5</td>
<td>58.1</td>
<td>63.5</td>
</tr>
<tr>
<td>Down</td>
<td>51.9</td>
<td>44.4</td>
<td>64.5</td>
<td>58.1</td>
<td>63.5</td>
</tr>
<tr>
<td>Cr1</td>
<td>91.2</td>
<td>84.8</td>
<td>77.4</td>
<td>61.1</td>
<td>89.2</td>
</tr>
<tr>
<td>MMRR</td>
<td>Cr2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>74.3</td>
<td>73.3</td>
<td>73.5</td>
<td>59.4</td>
<td>89.2</td>
</tr>
<tr>
<td>Down</td>
<td>73.3</td>
<td>74.3</td>
<td>73.5</td>
<td>59.4</td>
<td>89.2</td>
</tr>
<tr>
<td>Cr3</td>
<td>82.1</td>
<td>82.8</td>
<td>50.0</td>
<td>15.4</td>
<td>89.2</td>
</tr>
</tbody>
</table>
the candle chart, we can forecast the future candle chart form. We evaluated the forecast accuracy based on probability that forecasted direction is right, the root mean square error and the square correlation coefficient. The forecasting results described above indicate that the MMRR method has a better performance than 4-price method.

In this study, we applied the ARMA (or ARIMA) model with the ARCH error. In order to improve the forecast accuracy, however, consideration of General autoregressive conditional heteroscedasticity model or hybrid model may be required.

References


