Article

Geometric Case Based Reasoning for Stock Market Prediction

Se-Hak Chun [1,*] and **Young-Woong Ko** [2]

[1] Department of Business Administration, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea

[2] Department of Software, Hallym University, 1 Hallymdaehak-gil, Chucheon, Gangwon 24252, Korea; yuko@hallym.ac.kr

* Correspondence: shchun@seoultech.ac.kr; Tel.: +82-2-970-6487; Fax: +82-2-973-1349

check for updates

Abstract: Case based reasoning is a knowledge discovery technique that uses similar past problems to solve current new problems. It has been applied to many tasks, including the prediction of temporal variables as well as learning techniques such as neural networks, genetic algorithms, decision trees, etc. This paper presents a geometric criterion for selecting similar cases that serve as an exemplar for the target. The proposed technique, called geometric Case Based Reasoning, uses a shape distance method that uses the number of sign changes of features for the target case, especially when extracting nearest neighbors. Thus, this method overcomes the limitation of conventional case-based reasoning in that it uses Euclidean distance and does not consider how nearest neighbors are similar to the target case in terms of changes between previous and current features in a time series. These concepts are investigated against the backdrop of a practical application involving the prediction of a stock market index. The results show that the proposed technique is significantly better than the random walk model at $p < 0.01$. However, it was not significantly better than the conventional CBR model in the hit rate measure and did not surpass the conventional CBR in the mean absolute percentage error.

Keywords: artificial intelligence; case based reasoning; learning technique; knowledge discovery; data mining; financial prediction

1. Introduction

Traditionally, the prediction of stock markets has relied on statistical methods including multivariate statistical methods, autoregressive integrated moving average (ARIMA), and autoregressive conditional heteroscedasticity (ARCH) models. Recently, deep learning and other knowledge techniques have been extensively applied to the task of predicting financial variables. Case-based reasoning (CBR) is one of the most popular methodologies in knowledge-based systems. CBR solves a new problem by recalling and reusing specific knowledge from past experiences [1]. Research on stock prediction using the CBR technique is lacking, but CBR is used for both research and practical applications in many areas such as medical diagnosis, recommendation systems, cybersecurity detection, and others.

This paper shows how CBR can be applied to stock prediction using a geometric criterion for selecting similar cases to serve as an exemplar for the target. The proposed technique uses a shape distance method that uses the number of sign changes of features for the target case, especially when extracting nearest neighbors. Thus, this method overcomes the limitation of the conventional CBR in that it uses Euclidean distance and does not consider how nearest neighbors are similar to the target case in terms of changes between previous and current features in a time series. We investigated

these concepts against the backdrop of a practical application involving the prediction of a stock market index.

The rest of this paper is organized into five sections. Section 2 reviews CBR as a knowledge discovery technique. Section 3 introduces the proposed technique, which is called geometric CBR. Section 4 presents the case study. Section 5 discusses the results of the study. Finally, the concluding remarks are presented in Section 6.

2. Literature Review

2.1. Case-Based Reasoning

Case-based reasoning (CBR) is an approach for solving a new problem by remembering a previous similar situation and reusing information from, and knowledge of, that situation [2]. This concept assumes that similar problems have similar solutions, so CBR is an appropriate method for a practical domain focused on using real cases rather than rules or knowledge to solve problems. A general CBR cycle was described by the following four processes by Aamodt and Plaza [2]:

1. RETRIEVE the most similar case or cases.
2. REUSE the information and knowledge in that case to solve the problem.
3. REVISE the proposed solution.
4. RETAIN the parts of this experience likely to be useful for future problem solving.

According to this process, CBR solves a problem by retrieving one or more previous cases, reusing them to solve the problem, revising the potential solution based on the previous cases, and retaining the new experience by incorporating it into the existing case-base [2].

Conventional methods of prediction based on discrete logic usually seek the single best instance, or a weighted combination of a small number of neighbors in the observational space. An intelligent learning algorithm should therefore consider a "virtual" or composite neighbor whose parameters are defined by some weighted combination of actual neighbors in the case set. In this way, the algorithm can use the knowledge reflected in a larger subset of the case set rather than the immediate collection of proximal neighbors [3–11]. The procedure for case reasoning using composite neighbors and the Euclidean distance method is presented in Figure 1.

Step1. Begin with current case $x(t)$.
Step2. Seek the J neighboring cases $x(t_i)$ in the past which are closest to $x(t)$ according to the distance function: $d_i \equiv [x(t_i), \ x(t)]$.
Step3. Compute the sum of weights: $d_T = \sum_{i=1}^{J} d_i$.
Step4. Determine the relative weight of ith neighbor: $w_i = \frac{1}{J-1}\left[1 - \frac{d_i}{d_T}\right]$.
Step5. Find the successor $x(t_i \ + \ 1)$ of each case $x(t_i)$ in the set of neighbors.
Step6. Calculate the forecast for $t \ + \ 1$ as the weighted sum of successors: $\hat{x}(t \ + \ 1) = \sum_{i=1}^{J} w_i \, x(t_i \ + \ 1)$.

Figure 1. Predictive procedure using a numerical distance.

One of the issues with using a conventional CBR is that many previously experienced cases must be retrieved. The conventional CBR technique tends to retrieve a fixed number of neighbors in observational space. Thus, it always selects the same number of neighbors irrespective of an optimal number of similar neighbors according to target cases. This fixed number of neighbors raises a problem when some target cases should consider more similar cases while others should consider fewer ones. A problem occurs with conventional CBR when there are too many cases equidistant to target cases. Thus, it does not guarantee optimal similar neighbors for various target cases, which leads to the weakness of lowering predictability due to deviation from desired similar neighbors.

Chun and Park [8] suggested a model to find the optimal neighbors for each target case dynamically. Park et al. [10] suggested a new case extraction technique called statistical case-based reasoning (SCBR), which dynamically adapts the optimal number of neighbors by considering the distribution of distances between potential similar neighbors for each target case.

Conventional CBR using Euclidean distance does not consider how nearest neighbors are similar to the target case in terms of changes between previous and current features in time series. This paper proposes a new similarity measure called shape distance which compares how rise and fall signs between a target case and possible neighbors are similar to each other.

2.2. Case-Based Reasoning in Financial Area

CBR has been intensively exploited to solve problems in the domain of finance, including the predictions of corporate bond ratings [12–14], business failure [15–18], financial distress [19–21], bankruptcy [22–26], financial risk [27], real estate valuation [28], cost estimation [29], and prediction of the stock market [6,8,9,14].

Although recent research on stock prediction using the CBR technique is scarce, CBR is used in many areas such as medicine [30–36], recommendation systems [37–39], and cybersecurity [40,41].

3. Methodology

In this paper, we propose a shape distance method that selects nearest neighbors according to a slope similarity between two cases. In financial forecasting, hit rate in stock prediction may be an important decision tool for those who invest money in the stock market. Conventional CBR has focused on a numeric distance method and does not consider how possible cases can be similar to the target case in terms of features' shapes. We developed Kim and Kang's [42] distance measurement method for retrieving neighbors. Before explaining shape similarity, we first introduce the concept of numeric distances, which are used in conventional CBR methods.

3.1. Numeric Distances

The numeric distance refers to the conventional approach for determining the distance between two cases. Numeric distance is based on the differences of multiple features (or instances) between the target case and other learning cases in a stored case box. If one scalar target case x_i has m instances (or features), $m_i = \{a_{1i}, a_{2i}, a_{3i}, ..., a_{mi}\}$; then, the distance between x_i and x_j is given by Equation (1):

$$d_f \equiv (a_{1i} - a_{1j})^2 + \ldots + (a_{mi} - a_{mj})^2 \equiv \sum_{k=1}^{m} (a_{ki} - a_{kj})^2 \tag{1}$$

More specifically, if one vector target case X_i has n consecutive time series data and each consecutive datum has m features (or instances), then the feature distance between two cases x_i and x_j is given by Equation (2):

$$D_f \equiv (x_{i-n+1}^T - x_{j-n+1}^T)^2 + \ldots + (x_i^T - x_j^T)^2 \equiv \sum_{k=0}^{n-1} (x_{i-k} - x_{j-k})^2 \tag{2}$$

Distance metrics can be used, such as Manhattan distance or Gaussian distance. When the metric of choice is the standard Manhattan distance, the previous relationship becomes Equation (3):

$$D_f = \sum_{k=0}^{n-1} |x_{i-k} - x_{j-k}| \tag{3}$$

In this paper, we use one vector target case x_i, which has one day of stock information and has 4 features (or instances) such as open, high, low, and closed stock prices.

3.2. Shape Distance Method

Time series data may be characterized by their pattern of behavior in terms of rises and falls. Thus, when selecting neighbors, the slopes corresponding to two regional trajectories may be compared

to assess the similarity between two cases. Signs such as rises and falls between previous and current stock prices can be a simple shape distance. To understand this new distance method, consider a scalar variable such as the price of a stock. For example, let x_t denote the price of a stock at specific time t and m instances (or features), $m_t = \{a_{1t}, a_{2t}, a_{3t}, \dots, a_{mt}\}$. If we consider n consecutive time series data and each consecutive datum has m features (or instances), $x_t \equiv (x_{t-n+1}, \dots, x_{t-1}, x_t)^T$. Thus, we have a sequence of s consecutive time series data, $S_t \equiv (\text{sign}_{t-s+1}(s_{t-s}{}^* s_{t-s+1}), \dots, \text{sign}_t (x_{t-1}{}^* x_t))^T$. Let s_t denote the sign for observation x_t from the previous value x_{t-1}. We denote a difference in successive observations as follows (Equation (4)):

$$s_i = x_i - x_{i-1} \tag{4}$$

If there are n cases in a stored case box and if x_t is a target case, then the shape distance between two cases x_t and x_n is defined as follows (Equation (5)):

$$d_i = n - \sum_{k=1}^{n} U(S_t * S_{t-k}) \tag{5}$$

where U denotes the unit step function. The binary function U equals 1 if S_t. and S_{t-k} are both positive or both negative, and zero otherwise. When no matches in sign occur, the shape distance attains a maximum of $d_i = n$. Conversely, the shape distance reaches its minimum of $d_i = 0$ when all n pairs of slopes match. For simple analysis, we consider a one-day case as a target case. Thus, x_t has one day of information (x_t) and has only 4 instances: open, high, low, and closed stock prices. Figure 2 presents the procedure for selecting nearest neighbors using shape distance.

---

1. Perform exploratory data analysis (EDA): identify overall patterns and outliers.
2. Transform data for comparability (for example, differencing).
3. Perform a CBR machine.
    a. Obtain a shape information (fall or rise) from comparing a target case (t) with previous time series data (t-1).
    b. Seek the j neighboring cases x_k in the past which are closest to x_t according to the distance function: $d_i \equiv n - \sum_{k=1}^{n} U(S_t * S_{t-k})$.
    c. Compute the sum of weights: $d_T = \sum_{s=1}^{j} d_s$.
    d. Determine the relative weight of ith neighbor: $w_i = \frac{1}{j-1}\left(1 - \frac{d_i}{d_T}\right)$.
    e. Find the successor x_{k+1} of each case x_k in the set of neighbors.
    f. Calculate the forecast for x_{t+1} as the weighted sum of successors: $\hat{x}(t+1) = \sum_{i=1}^{j} w_i x(t_i + 1)$.
    g. Do loop from a to f until a final period of prediction in the test phase.
4. End Case Based Reasoning.
    → Calculate result from the model.
5. Detransform the variables.
    → Analyze results.

Figure 2. Predictive procedure using a shape distance method.

4. Application to Stock Market Data

4.1. The Data

With this case study, we aimed to investigate the effect of the proposed technique's predictive performance in forecasting a stock market index. The case study involved the prediction of the Dow Jones Industrial Average Index (DJI) for two experiments. The first experiment used a large data set for a learning phase which includes daily values starting from 29 January 1985. This is the largest possible dataset that can be obtained using Python from Yahoo Finance. The corresponding Python library is web.DataReader ('^DJI', 'yahoo', start='1985–01–29', end='2020–03–20'). Therefore, for the first case

study, the learning phase consisted of 8551 observations from 29 January 1985 to 31 December 2018, and the testing phase consisted of 307 observations from 2 January 2019 to 20 March 2020. The second experiment dealt with a relatively recent dataset beginning from 2006. The learning phase consisted of 3271 observations from 3 January 2006 to 31 December 2018, and the testing phase consisted of 307 observations from 2 January 2019 to 20 March 2020. The raw variables for DJI prediction were as follows:

Opening value (Open). The value of the Dow Jones Industrial Average Index (DJI) at the beginning of the trading day.
*Daily high (High).* The high value for the DJI.
Daily low (Low). The low value for the DJI.
Daily close (Close). The close value for the DJI.

4.2. Model Construction

An exploratory plot of the Dow Jones Industrial Average Index (DJI) is given in Figure 3. Other exploratory plots for the raw data series are shown in Figures 4–6. Figure 4 depicts the trajectory of the opening value of the DJI. Figure 5 plots the high value of the DJI, and Figure 6 displays the low value of the DJI.
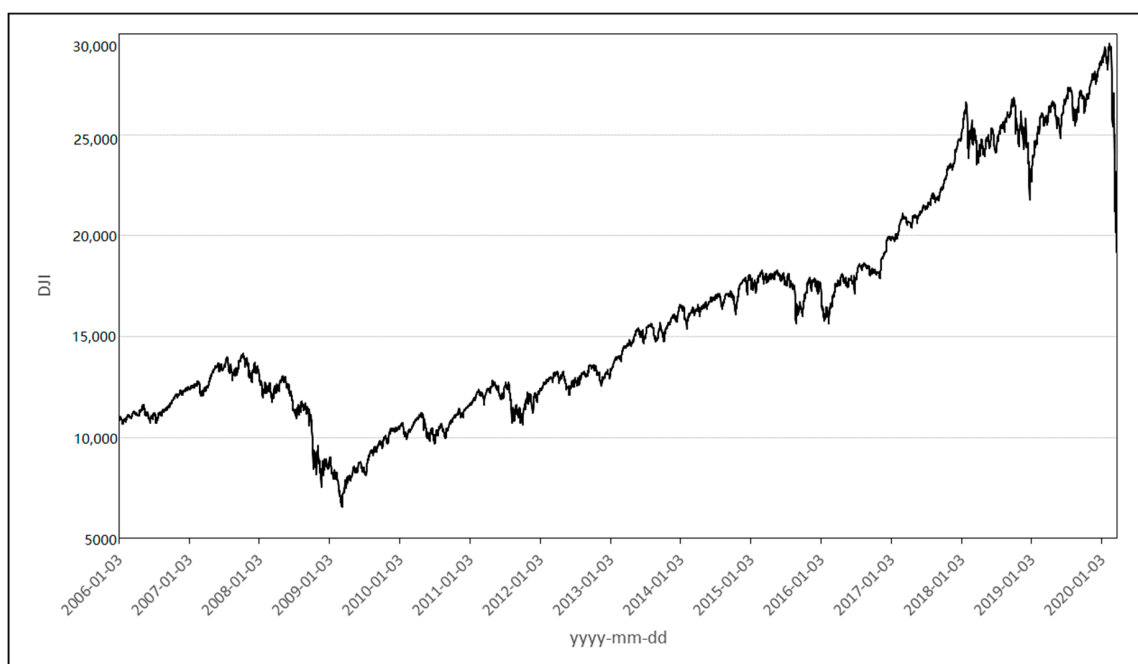
Figure 3. History of the Dow Jones Industrial Average Index (DJI) from 3 January 2006 to 20 March 2020.

In constructing the predictive model for the DJI, the input variables were first transformed. For financial variables, stationarity can often be obtained through a logarithmic and a differencing operation [9]. However, because the data we used in the case study did not need to eliminate the effects of measurement units among variables, only a differencing procedure was performed. For example, opening value at t time ($Open_t$) was transformed to be $dOpen_t$ ($Open_t$ - $Open_{t-1}$) through the differencing procedure. Other input variables such as $High_t$, Low_t, and $Close_t$ were transformed to be $dHigh_t$, $dLow_t$, and $dClose_t$, respectively. As shown in Figure 7, these variables were used for the prediction engine of geometric CBR to produce the predicted value of $dClose_t$. Then, finally, the predicted value of the closing price at t + 1 ($pClose_{t+1}$) was obtained from a de-transforming procedure by adding the predicted value of $dClose_t$ to the previous actual closing price at t ($Close_t$). Figure 7 presents an overview of pre-processing and post-processing for producing prediction values by geometric CBR.
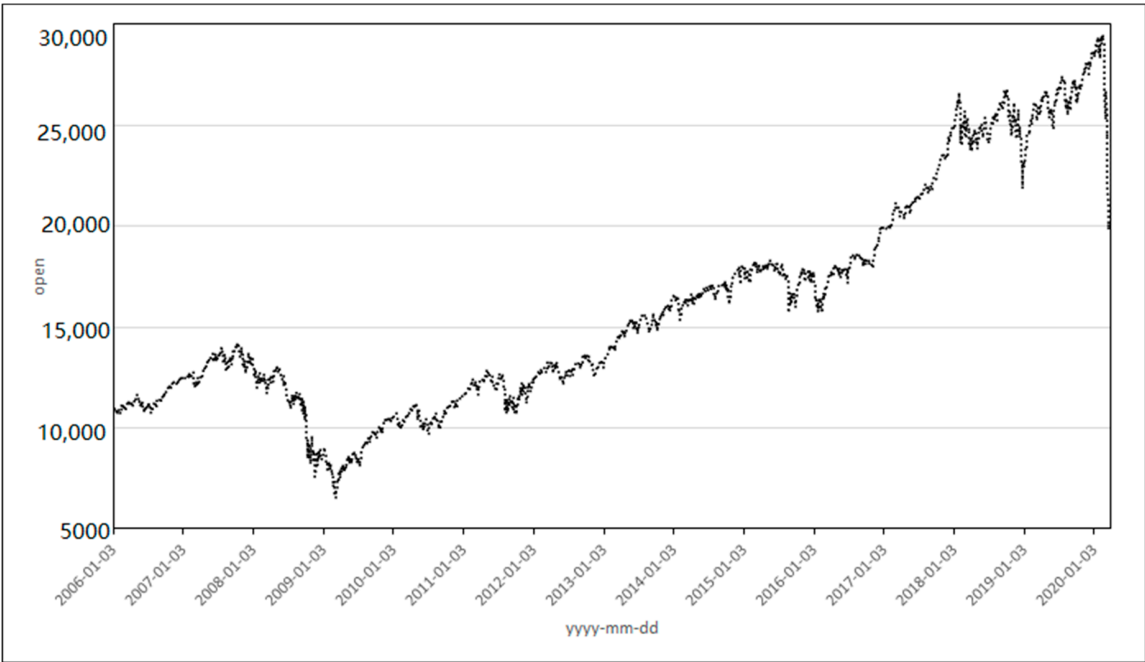
Figure 4. Time series for the opening value for the Dow Jones Industrial Average Index (DJI) from 3 January 2006 to 20 March 2020.
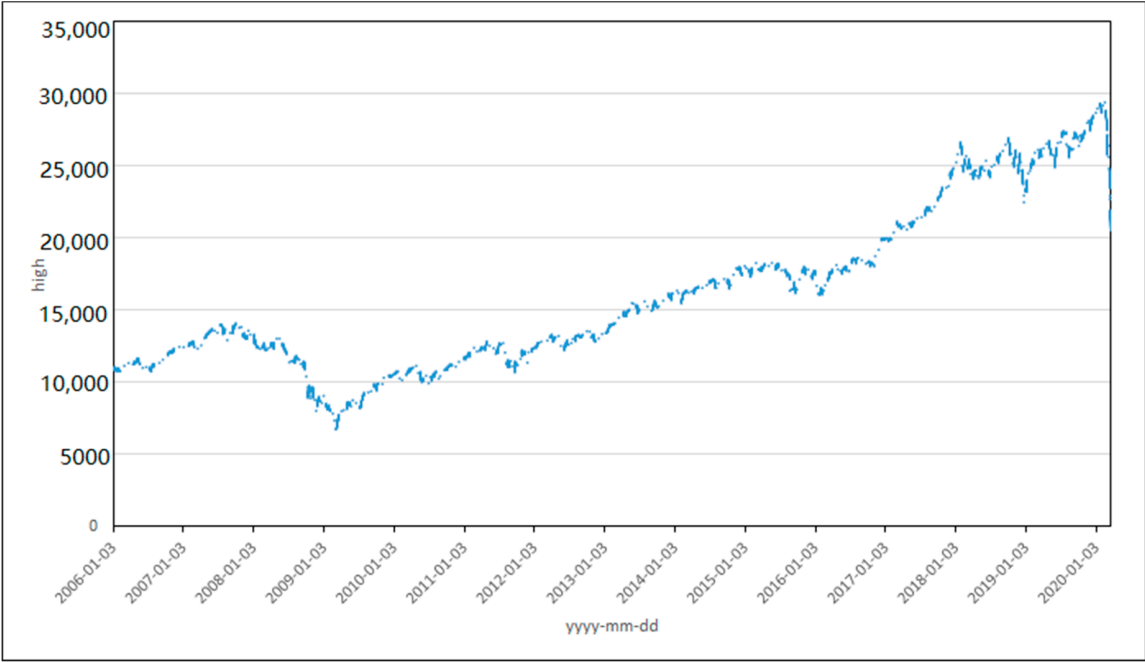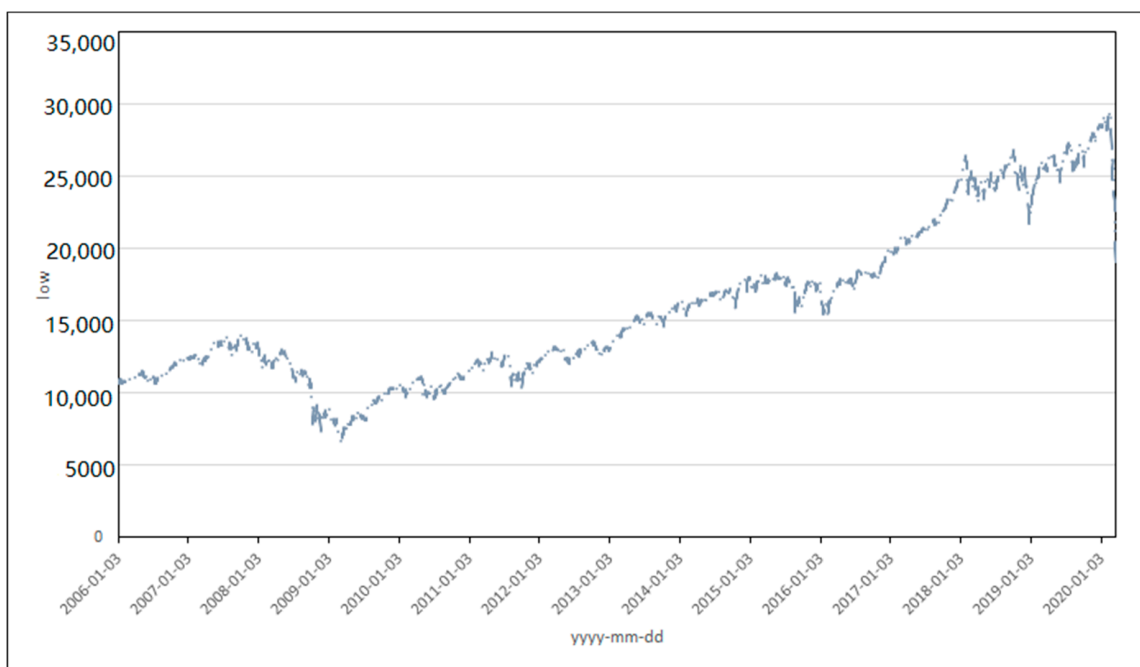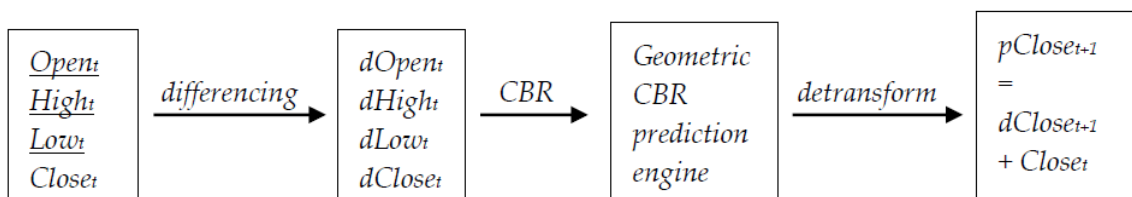

Figure 5. Time series for the high value for the Dow Jones Industrial Average Index (DJI) from 3 January 2006 to 20 March 2020.

Figure 6. Time series for the low value for the Dow Jones Industrial Average Index (DJI) from 3 January 2006 to 20 March 2020.

Figure 7. Overview of preprocessing and postprocessing for geometric CBR prediction.

5. Results

For much of this century, the random walk model of stock prices has served as a pillar of accepted wisdom in financial economics. One implication of the random walk model is that obvious patterns in the economy are already incorporated in the valuation of stock prices and financial markets. This is the rationale behind technical analysis in forecasting stock prices based solely on variables pertaining to the market. The performance results among the predictive models such as the random walk (RW), a conventional CBR using the Euclidean distance method, and geometric CBR (GCBR) using a shape distance method are presented in Tables 1 and 2.

Table 1 summarizes the hit rates (the proportion of correct forecasts) in the test data. Hit rates show how well the GCBR predicted the direction of price changes for the closing prices of the DJI. Table 1 indicates that GCBR seems to be significantly more accurate than the random walk model at $p < 0.01$. We also tested the null hypothesis H_0, which states that the proposed GCBR does not produce more accurate performance than the conventional CBR method. Thus, the decision of *Reject H_0* in Table 1 represents the superior performance of GCBR over the conventional CBR.

When the dataset is small and relatively recent (2006+), GCBR is more accurate than CBR overall. However, the results showed that the outperformance of GCBR is statistically significant only for the case when there is one nearest neighbor. For all other experiments, GCBR was not significantly better than the conventional CBR method. However, when the dataset was large (1985+), GCBR did not perform better than the conventional CBR method. This is because when predicting with few similar incidents in the past, as the dataset becomes larger and when the distance is equal, the possibility

increases of using older cases that may have lost their relevance. The results of the first experiment (1985+) implied that when the number of similar cases was in the range of 30 to 100, GCBR exhibited superior performance to CBR. With 75 nearest neighbors, GCBR performed better than CBR to a statistically significant degree, with $p < 0.1$.

Table 1. Hit rate among the forecasting models for the test phase.

Model Neighbors	Random Walk	GCBR vs. CBR (1985+) *	GCBR vs. CBR (2006+) *	z-Value ** (p-Value)	Decision
1		50.1% vs. 52.11%	53.7% vs. 45.2%	2.099 (0.036)	Reject H_0
2		46.2% vs. 50.8%	52.4% vs. 52.7%	−0.080 (0.936)	Accept H_0
3		51.1% vs. 55.4%	48.9% vs. 51.8%	−0.645 (0.522)	Accept H_0
5		48.2% vs. 49.8%	50.8% vs. 50.1%	0.161 (0.872)	Accept H_0
10		46.2% vs. 50.1%	50.1% vs. 52.1%	−0.484 (0.631)	Accept H_0
20		52.1% vs. 53.0%	53.0% vs. 51.1%	0.484 (0.628)	Accept H_0
30	46.2%	54.1% vs. 53.7%	52.8% vs. 51.1%	0.403 (0.687)	Accept H_0
50		52.8% vs. 52.1%	55.7% vs. 51.1%	1.132 (0.258)	Accept H_0
75		57.7% vs. 50.8%	53.1% vs. 53.4%	1.728 (0.083) ***	Reject H_0
100		54.2% vs. 50.8%	50.4% vs. 53.7%	−0.888 (0.379)	Accept H_0
150		47.6% vs. 52.8%	52.8% vs. 51.8%	0.242 (0.809)	Accept H_0
200		50.8% vs 53.7%	54.3% vs. 49.5%	0.970 (0.332)	Accept H_0
300		50.8% vs 52.4%	55.0% vs. 52.4%	0.647 (0.518)	Accept H_0

* 1985+ indicates the results of the first experiment, of which the learning phase began 29 January 1985. * 2006+ indicates the results of the second experiment, of which the learning phase began 3 January 2006. ** The z-value and p-values are based on the 2006+ dataset. A z-test for differences in two proportions was used to determine whether there was a difference between two population proportions, which is defined as $Z \cong \frac{(p_1-p_2)-(\pi_1-\pi_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$. where p_i are sample proportions, π_i are population proportions, n_i are sample sizes for groups, and $\bar{p}$ is a pooled estimate of the proportion of success in a sample of both groups, $\bar{p} = (n_1 p_1 + n_2 p_2)/(n_1 + n_2)$. *** This signifies the z- and t-values for the 1985+ dataset when the number of nearest neighbors was 75.

Table 2. Mean Absolute Percentage Error (MAPE) * and pairwise ** t-tests.

Models Neighbors	Random Walk	GCBR vs. CBR (1985+)	GCBR vs. CBR (2006+)	z-Value (p-Value)	Decision
1		0.934 vs. 0.979	0.941 vs. 1.032	−0.784(0.433)	Accept H_0
2		0.894 vs. 0.914	0.917 vs. 0.942	−0.206(0.836)	Accept H_0
3		0.882 vs. 0.889	0.906 vs. 0.905	0.001(0.998)	Accept H_0
5		0.875 vs. 0.869	0.884 vs. 0.888	−0.032(0.974)	Accept H_0
10		0.870 vs. 0.870	0.870 vs. 0.872	−0.018(0.985)	Accept H_0
20		0.859 vs. 0.859	0.864 vs. 0.859	0.046(0.963)	Accept H_0
30	0.856	0.857 vs. 0.858	0.860 vs. 0.858	0.015(0.987)	Accept H_0
50		0.857 vs. 0.854	0.854 vs. 0.855	−0.009(0.992)	Accept H_0
75		0.858 vs. 0.854	0.856 vs. 0.854	0.014(0.988)	Accept H_0
100		0.857 vs. 0.854	0.857 vs. 0.854	0.029(0.976)	Accept H_0
150		0.858 vs.0.854	0.856 vs. 0.856	−0.003(0.997)	Accept H_0
200		0.857 vs. 0.854	0.855 vs. 0.857	−0.020(0.983)	Accept H_0
300		0.856 vs. 0.856	0.853 vs. 0.857	−0.028(0.976)	Accept H_0

* The MAPE is a commonly used statistical measure of goodness of fit in quantitative forecasting methods because it produces a measure of relative overall fit. Mean Absolute Percent Error (MAPE) is calculated by averaging the percentage difference between the fitted (forecast) line and the original data. MAPE is defined as $\frac{\sum \left|\frac{y_t-f_t}{y_t}\right| \times 100}{n}$, where y represents the original series, f the forecast, and n the number of observations. ** Pairwise t-tests of the predictive models for the test phase. The comparison is based on the MAPE of the residuals.

In the second experiment (2006+), the performance of GCBR was most accurate when the number of nearest neighbors was 50; the hit rate of GCBR being 0.5570 and the hit rate of the random walk model being 0.4625. This result implied the possibility of GCBR being superior to a conventional CBR in terms of hit rate.

CBR performed better with the larger dataset (1985+) as it implements a numerical distance method in the learning phase. For GCBR, the first experiment implies that to enhance performance, the number of nearest neighbors must reach a certain level, in this case, 30. This further implies that when two relevant cases found through CBR have the same distance, optimizing using the most recent case will enhance performance.

Table 2 presents the results of MAPE and of the *t*-test for the difference in performance of the random walk, conventional CBR and GCBR methods, respectively. The GCBR model does not seem to surpass other models. When the number of neighbors was 300, the MAPE of GCBR was best at 0.853. With a large dataset, there are several models where the performance of GCBR was better than CBR, but the difference was insignificant. The two methods also did not exhibit significant performance differences in terms of MAPE. We originally intended to optimize the GCBR model using hit rates; thus, MAPE was expected to be lower. Such results further implied that shape-distance-based GCBR has the potential to enhance MAPE-based performance.

6. Concluding Remarks and Future Work

This paper proposed a shape distance method for selecting nearest neighbors in case-based reasoning. Concepts were investigated against the backdrop of a practical application involving the prediction of a stock market index. The results of the case study are summarized as follows:

- The proposed technique, GCBR, is significantly better than the random walk model at $p < 0.01$.
- Overall, GCBR is more accurate than conventional CBR models in terms of hit rate. However, the superiority was not statistically significant compared to conventional CBR models.
- The GCBR was not found to surpass a conventional CBR in terms of MAPE.
- GCBR outperformed a conventional CBR in terms of MAPE when the number of nearest neighbors was small and the dataset was recent and smaller.
- When the dataset was larger, GCBR performed significantly more accurately than CBR when the number of nearest neighbors was 75.
- The proposed method has the possibility to improve predictability. Thus, in future research, we propose implementing the shape distance method along with consecutive time series data in searching the nearest neighbors, which would improve GCBR though validating the predictability of GCBR. A promising direction for the future would involve finding optimal neighbors by combining the numeric distance and shape distance methods.

Author Contributions: Conceptualization, S.-H.C. and Y.-W.K.; methodology, S.-H.C.; software, S.-H.C. and Y.-W.K.; validation, S.-H.C. and Y.-W.K.; formal analysis, S.-H.C. and Y.-W.K.; investigation, Y.-W.K.; data curation, Y.-W.K.; visualization, Y.-W.K.; funding acquisition, S.-H.C. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ince, H. Short term stock selection with case-based reasoning technique. *Appl. Soft Comput.* **2014**, *22*, 205–212. [CrossRef]
2. Aamodt, A.; Plaza, E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Commun. Eur. J. Artif. Intell.* **1994**, *7*, 39–59. [CrossRef]
3. Kim, S.H.; Chun, S.H. Graded forecasting using an array of bipolar predictions: Application of probabilistic neural networks to a stock market index. *Int. J. Forecast.* **1998**, *14*, 323–337. [CrossRef]
4. Chun, S.-H.; Kim, K.; Kim, S.H. Chaotic analysis of predictability versus knowledge discovery techniques: Case study of the Polish stock market. *Expert Syst.* **2002**, *19*, 264–272. [CrossRef]

5. Chun, S.-H.; Kim, S.H. Impact of momentum bias on forecasting through knowledge discovery techniques in the foreign exchange market. *Expert Syst. Appl.* **2003**, *24*, 115–122. [CrossRef]

6. Chun, S.-H.; Kim, S.H. Data mining for financial prediction and trading: Application to single and multiple markets. *Expert Syst. Appl.* **2004**, *26*, 131–139. [CrossRef]

7. Chun, S.-H.; Kim, S.H. Automated generation of new knowledge to support managerial decision making: Case study in forecasting a stock market. *Expert Syst.* **2004**, *21*, 192–207. [CrossRef]

8. Chun, S.-H.; Park, Y.-J. Dynamic adaptive ensemble case-based reasoning: Application to stock market prediction. *Expert Syst. Appl.* **2005**, *28*, 435–443. [CrossRef]

9. Chun, S.-H.; Park, Y.-J. A new hybrid data mining technique using a regression case based reasoning: Application to financial forecasting. *Expert Syst. Appl.* **2006**, *31*, 329–336. [CrossRef]

10. Park, Y.J.; Kim, B.C.; Chun, S.H. New knowledge extraction technique using probability for case-based reasoning: Application to medical diagnosis. *Expert Syst.* **2006**, *23*, 2–20. [CrossRef]

11. Chun, S.-C.; Kim, J.; Hahm, K.-B.; Park, Y.-J.; Chun, S.-H. Data mining technique for medical informatics: Detecting gastric cancer using case-based reasoning and single nucleotide polymorphisms. *Expert Syst.* **2008**, *25*, 163–172. [CrossRef]

12. Dutta, S.; Shekkar, S. Bond rating: A non-conservative application of neural networks. *Int. Jt. Conf. Neural Netw.* **1988**, *2*, 443–450.

13. Shin, K.S.; Han, I.A. Case-based approach using inductive indexing for corporate bond rating. *Decis. Support Syst.* **2001**, *32*, 41–52. [CrossRef]

14. Kim, K.; Han, I. Maintaining case-based reasoning systems using a genetic algorithms approach. *Expert Syst. Appl.* **2001**, *21*, 139–145. [CrossRef]

15. Li, H.; Sun, J. Gaussian case-based reasoning for business failure prediction with empirical data in China. *Inf. Sci.* **2009**, *179*, 89–108. [CrossRef]

16. Li, H.; Huang, H.-B.; Sun, J.; Lin, C. On sensitivity of case-based reasoning to optimal feature subsets in business failure prediction. *Expert Syst. Appl.* **2010**, *37*, 4811–4821. [CrossRef]

17. Li, H.; Sun, J. Predicting business failure using an RSF-based case-based reasoning ensemble forecasting method. *J. Forecast.* **2013**, *32*, 180–192. [CrossRef]

18. Yip, A.Y.N. Business failure prediction: A case-based reasoning approach. *Rev. Pac. Basin Financ. Mark. Policies* **2006**, *9*, 491–508. [CrossRef]

19. Li, H.; Sun, J. Ranking-order case-based reasoning for financial distress prediction. *J. Knowl. Based Syst.* **2008**, *21*, 868–878. [CrossRef]

20. Li, H.; Sun, J.; Sun, B.-L. Financial distress prediction based on OR-CBR in the principle of k-nearest neighbors. *Expert Syst. Appl.* **2009**, *36*, 643–659. [CrossRef]

21. Li, H.; Sun, J. On performance of case-based reasoning in Chinese business failure prediction from sensitivity, specificity, positive and negative values. *J. Appl. Soft Comput.* **2011**, *11*, 460–467. [CrossRef]

22. Bryant, S.M. A case-based reasoning approach to bankruptcy prediction modeling. *Intell. Syst. Account. Financ. Manag.* **1997**, *6*, 195–214. [CrossRef]

23. Elhadi, M.T. Bankruptcy support system: Taking advantage of information retrieval and case-based reasoning. *Expert Syst. Appl.* **2000**, *18*, 215–219. [CrossRef]

24. Park, C.S.; Han, I. A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction. *Expert Syst. Appl.* **2002**, *23*, 255–264. [CrossRef]

25. Ahn, H.; Kim, K.J. Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach. *Appl. Soft Comput.* **2009**, *9*, 599–607. [CrossRef]

26. Yip, A.Y.N. Predicting business failure with a case-based reasoning approach. In Proceedings of the Knowledge-Based Intelligent Information and Engineering Systems: 8th International Conference, Wellington, New Zealand, 20–25 September 2004.

27. Kapdan, F.; Aktas, M.G.; Aktas, M.S. Financial risk prediction based on case based reasoning methodology. In Proceedings of the 2019 Innovations in Intelligent Systems and Applications Conference (ASYU) Intelligent Systems and Applications Conference (ASYU), Izmir, Turkey, 31 October–2 November 2019.

28. Yeh, I.-C.; Hsu, T.-K. Building real estate valuation models with comparative approach through case-based reasoning. *Appl. Soft Comput.* **2018**, *65*, 260–271. [CrossRef]

29. Relich, M.; Pawlewski, P. A case-based reasoning approach to cost estimation of new product development. *Neurocomputing* **2018**, *272*, 40–45. [CrossRef]

30. Lei, Z.; Yin, D. Intelligent generation technology of sub-health diagnosis case based on case reasoning. In Proceedings of the 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 11–13 October 2019.

31. Elisabet, D.; Sensuse, D.I.; Al Hakim, S. Implementation of case-method cycle for case-based reasoning in human medical health: A systematic review. In Proceedings of the 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 29–30 October 2019.

32. Chowdhury, A.R.; Banerjee, S. Case Based reasoning in retina abnormalities detection. In Proceedings of the 2019 4th International Conference on Information Systems and Computer Networks (ISCON) Information Systems and Computer Networks (ISCON), Mathura, India, 21–22 November 2019.

33. Lamy, J.-B.; Sekar, B.; Guezennec, G.; Bouaud, J.; Seroussi, B. Hierarchical visual case-based reasoning for supporting breast cancer therapy. In Proceedings of the 2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME), Tripoli, Lebanon, 17–19 October 2019.

34. Bentaiba-Lagrid, M.B.; Bouzar-Benlabiod, L.; Rubin, S.H.; Bouabana-Tebibel, T.; Hanini, M.R. A case-based reasoning system for supervised classification problems in the medical field. *Expert Syst. Appl.* **2020**, *150*, 113335. [CrossRef]

35. Lamy, J.-B.; Sekar, B.; Guezennec, G.; Bouaud, J.; Séroussi, B. Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach. *Artif. Intell. Med.* **2019**, *94*, 42–53. [CrossRef]

36. Silva, G.C.; Carvalho, E.E.O.; Caminhas, W.M. An artificial immune systems approach to case-based reasoning applied to fault detection and diagnosis. *Expert Syst. Appl.* **2020**, *140*, 112906. [CrossRef]

37. Sun, J.; Zhai, Y.; Zhao, Y.; Li, J.; Yan, N. Information acquisition and analysis technology of personalized recommendation system based on case-based reasoning for internet of things. In Proceedings of the 2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Zhengzhou, China, 18–20 October 2018.

38. Supic, H. Case-based reasoning model for personalized learning path recommendation in example-based learning activities. In Proceedings of the 2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), WETICE Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Paris, France, 27–29 June 2018.

39. Corrales, D.C.; Ledezma, A.; Corrales, J.C. A case-based reasoning system for recommendation of data cleaning algorithms in classification and regression tasks. *Appl. Soft Comput.* **2020**, *90*, 106180. [CrossRef]

40. Nunes, R.C.; Colomé, M.; Barcelos, F.A.; Garbin, M.; Paulus, G.B.; Silva, L.A.D.L. A case-based reasoning approach for the cybersecurity incident recording and resolution. *Int. J. Softw. Eng. Knowl. Eng.* **2019**, *29*, 1607–1627. [CrossRef]

41. Abutair, H.; Belghith, A.; AlAhmadi, S. CBR-PDS: A case-based reasoning phishing detection system. *J. Ambient. Intell. Humaniz. Comput.* **2019**, *10*, 2593–2606. [CrossRef]

42. Kim, S.H.; Kang, D.S. A new geometric criterion for case based reasoning in time series prediction. *Spring Conf. Proc. Korean Inst. Ind. Eng.* **1997**, *4*, 191–194.