

Towards the Ensemble: IPCBR Model in Investigating Financial Bubbles

Francis Ekpenyong, Georgios Samakovitis, Stelios Kapetanakis and Miltos Petridis

Abstract — Asset value predictability has always been a major research concern in financial market especially when considering the impact of unprecedented market fluctuations on the behavior of market participants.

This paper presents preliminary results toward the building a reliable forward problem on ensemble approach IPCBR model, that leverages the capabilities of Case based Reasoning(CBR) and Inverse Problem Techniques (IPTs) to describe and model abnormal stock market fluctuations (often associated with asset bubbles) using datasets from historical stock market prices. The framework uses a rich set of past observations and geometric pattern description and to come up with a model to generalize those patterns onto observations using PCA and k-means to formulate the forward problem.

This research work presents a formative strategy aimed to determine the causes of behavior, rather than predict future time series points which brings a novel perspective to the problem of asset bubbles predictability, and a deviation from the existing research trend. The results depicts the stock dynamics and statistical fluctuating evidence associated with the envisaged bubble problem.

Index Terms — Artificial Intelligence, Asset Bubble, Case-based Reasoning, Inverse Problems, Machine learning.

INTRODUCTION

Developments in Artificial Intelligence (AI) and Machine Learning [1]-[3] have revealed numerous spectacular outcomes in diverse fields. Financial markets is no exception, as it is one of the fields where AI solutions impact directly on real time implementation and achieve exceptional performance in a variety of complicated tasks.

In the wake of various equivocal concerns in asset value predictability, researchers have proposed various econometric techniques to consistently predict fluctuations in financial asset prices (often associated with asset bubbles) [4]-[6]. While these models promise improved predictive capability, they are yet to receive wider acceptance in practice.

The reason is not far the stochastic nature of asset fluctuations making it Difficult to build reliable models. A major concern arises with the increase in complexity of the problem, which makes it difficult to mathematically formulate the problems, and as such leads to the choice of parameters being set by heuristics. This introduces certain deficiencies in reliability and explainability, specifically

because it becomes very hard to identify which parameters need to be optimized and in what way, in order to improve the descriptive power of the model.

Based on the above, we propose an ensemble Inverse Problem Case Base Reasoning (IPCBR) model that uses the simplicity and applicability of CBR to deliver a more robust representation of asset value fluctuation patterns (and their subsequent classification as potential asset bubbles) and the successes of the Inverse Problem to identify the factors that most likely cause such patterns.

The outcome of this will then be used as a case base for standard Case-based Reasoning process and will be evaluated against some known episodic data retrieved from the Yahoo Finance and human expert advice. This paper is organized as follows: Section II outlines the class of asset bubble problems to be addressed in this research, and outlines the proposed relevant features/qualities of CBR that make it suitable, as well as the overall IP formulation approach. An insight to the Inverse problems is given in Section III with and its application areas, while Section IV provides an articulation of the overall model to be used where the Inverse Problem formulation component is discussed. Section V focuses on the implementation where the experimental outcome is presented and discussed. The paper closes with a critical discussion on the major contributions this work intends to deliver, and a set of relevant concluding observations.

II. RELATED WORK

The academic literature provides frequent reference to so-called asset bubbles, most often focusing on their theoretical or conceptual evolution and advancement. Bubbles are often defined relative to the fundamental value of an asset [7], [8]. While numerous theories exist on how bubbles form, one basic explanation may be that investors hold an asset even though the asset's price exceeds its fundamental value, with view to profiting from a later sale [9]. Detecting a bubble in real time is quite challenging, as what attributes to the fundamental value is difficult to pin down. Although every bubble differs in their initiation and specific details, they often share patterns that may be use to recognize their existence.

Numerous researchers have proposed methods for detecting asset bubbles [10]-[12]. While the existing literature mainly focuses on econometric prediction methods, machine learning algorithms have been adopted over the

Published on July 21, 2020.
Francis Ekpenyong, University of Brighton, UK.
(corresponding e-mail: F.Ekpenyong@brighton.ac.uk)
Georgios Samakovitis, University of Greenwich, UK.
(e-mail: G.Samakovitis@gre.ac.uk)

Stelios Kapetanakis, University of Brighton, UK.
(e-mail: S.Kapetanakisg@brighton.ac.uk).
Miltos Petridis, Middlesex University, UK.
(e-mail: M.Petridis@mdx.ac.uk)

years towards attempting to predict asset bubbles [13],[14]. However, while both academic and trade literature in Finance and Economics have long been examining the occurrence of asset bubbles, that literature falls well beyond the scope of this work. In this study, we adopt a relatively narrow definition of asset bubbles as patterns which can be described as, "a short-term continuous, sustained, and extra-ordinary price growth of an asset or portfolio of assets, which is followed by an equally extra-ordinary price decay in a comparably short period". The reader is referred to [15] for a full description of bubbles adopted in this research.

A. Case Based Reasoning process

Case-based Reasoning [16] is one of emerging field of Artificial intelligence research area. It is a paradigm for combining problem solving and learning, and has recently attracted research attention due to its simplicity and flexibility. This methodology was used in various application domain and it offers enormous advantages over other AI-based techniques in all the fields where experiential knowledge is readily available, like engineering [3], production planning [17], medicine [18] and other. Its operations is based on the concept proposed by [16] introducing four key phases: Retrieve, Reuse, Revise, and Retain as shown in figure 1.

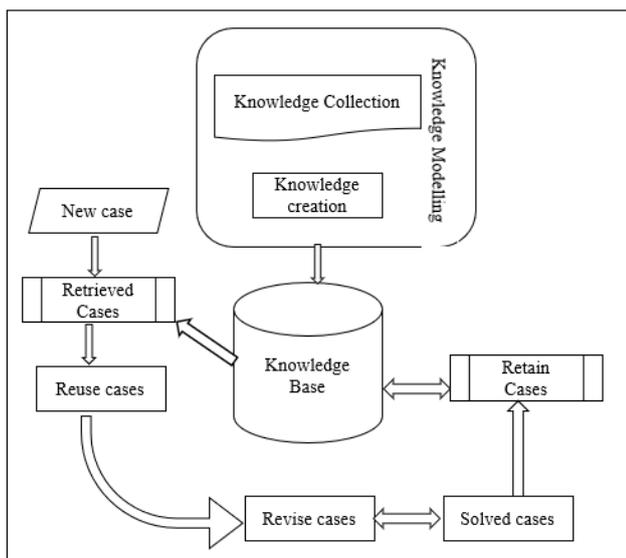


Fig. 1: Case-based Reasoning Cycle.

Retrieve- In terms of case retrieval, a new case characterized by a primary problem description is used to retrieve a case from the repository. The similarity measure computes the relationship between a new case and previous cases stored in the case base, which can be simple or more complex measure depending on the application domain and features used to define cases.

Reuse- Case reuse, also otherwise referred to as case adaptation, normally maps the solution from previous stored cases to the target problem, where the retrieved case is carefully selected and recommended as a candidate solution to the new problem. It then becomes immediate solution if it can be fittingly applied to the target case and could as well be used for the latter solution.

Revise- This step that maps previous solutions to the target

situation, in such a way that it can be applied to it. In which case, the previous cases need to be further scrutinized to ascertain whether they are appropriate for the current situation and whether they can solve the target problem. The solution is tested for success and repaired if failed. If revision is not possible, the system fails in finding a suitable solution to the target.

Retain- it evaluates the obtained solution and it decides whether to retain the new solved case in memory. This stage is a dynamic procedure of adding and removing cases seeking to enhance the efficiency of the CBR model. Useful knowledge is retained for future reuse, and the case base is updated by a new learned case, or by modification of some existing cases.

B. Case Definition and Representation

Case representation has posed some fundamental issues in Case-based Reasoning methodology due to existence of some unaccustomed features that do not exist or align with the processing of the traditional "attribute-value" data representation, and the direct manipulation of continuous, high dimensional data in the time series domain [19].

We addressed this issue by forming a library pattern of observations and treated every group as a case category. The adopted approach followed the concept proposed in [20], where the entire Time series is split into smaller sequences of patterns, by decomposing the series into a sequence of rolling observations, or rolling windows, in which case every observation in the pattern constitutes the case which may attain a predefined upward, steady or declining pattern as shown in figure 2



Fig. 2: Sample case patterns.

This also infers that an interval comprising a series of three observation patterns can be easily recognized as constituting a case. Further analysis and matching of all the similar cases using appropriately selected algorithm makes it possible to discover a specific relation to the pattern.

C. Case Retrieval

Case retrieval and the adaptation phases are the most challenging phases in the CBR cycle. The Retrieval phase which based on a chosen similarity measure involves a process where the CBR methodology matches a new case with archive of historical cases with the suggested solution plan and returns results of closest matches to be adapted. In this paper, we adopted the Clustering technique, an unsupervised machine learning method in constructing the forward problem [15], in which case, the retrieval process only considers cases that are in the same cluster as the new case.

D. Clustering methods

Clustering is known to be one of the basic concepts in pattern recognition. It is unsupervised classification through partitioning data into segments of homogeneous data objects based on similarity of some features and dissimilarity to the objects in other clusters. This technique is predominantly

used as a knowledge discovery tool in modern machine learning with applications in a wide range in sciences, marketing, geology, medicine, etc as well as various data mining tasks involving time series [21] and financial markets[22] [23] In this paper we consider the K-means technique for clustering, which usually takes the Euclidean distance between the feature and feature :

$$D(x, y) = \sum_i^n (x_i - y_i)^2 \quad (1)$$

where: $d(x; y)$ is the distance of data x to the centre of the cluster y , x_i is data position i in n data, and y_j is data position j in n data.

In other words, Given a set of observations $X_1, X_2, X_3, \dots, X_n$, K-Means clustering aims to partition the N observations into ($K \leq N$) sets $S = \{S_1, S_2, S_3, \dots, S_n\}$ with the objective to minimize the within-cluster sum of squares (WCSS) given by:

$$E = \sum_{j=1}^k \sum_{i=1}^n \|x_i^j - C_j\|^2 \quad (2)$$

where k represents the number of clusters formed, n is the number of cases, C_j denotes the centroid of i -th cluster, $\|(\)\|^2$ is the function which produces squared distance between two given points.

E. Algorithmic Approach to Case Matching in bubble

Consequently, the concept of the CBR proposed by [16], [24] and [25] is aligned with the task of pattern recognition with the algorithm illustrated in Fig. 3.

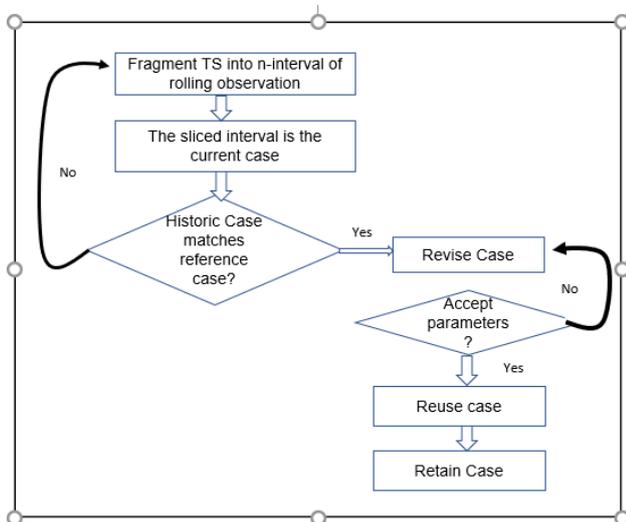


Fig. 3: Case Matching in bubble structure.

Firstly, a fragment of the series is created and divided into n -interval patterns consisting of sequential rolling observation which are being stored as historical cases of n -interval patterns. This sliced interval in the series of a new observation now makes the current case to be referenced during the retrieval process. All cases satisfying the pattern matching conditions are retrieved. If there is no perfect match as the case may be, the closest cases are retrieved, otherwise the fragmentation process is done again with the new case. The retrieved cases would then be revised and tested against the stated parameters, if satisfied, then the case is then reused

and the solution Retained in the database.

III INVERSE Problem

Since its first appearance in the 1960s, the term "inverse problem", has significantly changed from what was notably to designate in geophysics (i.e. the determination, through input/output or cause-effect experiments of unknowns in the equations), to a contemporary "inverse problem" that designates the best possible reconstruction of missing information, in order to estimate either the identification of sources or of the cause, or the value of undetermined parameters [26]. Inverse problem involves mapping between objects of interest (parameters), and acquired information about these objects, (data or measurements) [27].

This technique, otherwise referred to as model inversion [28] exists in many fields, usually in a situation where one is interested in finding a model that typically approximates observational data. Any inverse theory requirement is to relate a physical parameter, say "u" that describes a model to acquire observations making up some set of data, say 'f'. Assuming the underlying concept of the model has a clear and well-stated representation, then an operator can be assigned a relation or mapping u to f through the equation:

$$F = Ku \quad (3)$$

where f is an N - dimensional constant coefficients data vector and u is an M -dimensional model parameter, and K (the Kernel) is an $N \times M$ matrix containing only constant coefficients.

In a case where the operator f is linear, the inverse problem is termed as linear and the direct inverse is easy to find; otherwise it is a non-linear inverse problem, and termed ill-posed problem which poses considerable difficulties in solving.

A. Methods for solving the Inverse Problems

Inverse problems theory involves a combination of mathematical techniques that operate on reduced volume of data in a problem, with a primary objective of obtaining useful information relevant to the real physical system in question. Many inverse problems exist; [26] classified various approaches to solving inverse problems (in line with the basis searched for) into three main categories namely, (i) Regularization of Ill-Posed Problems, (ii) Stochastic or Bayesian Inversion [29], and (iii) Functional analysis, a decision making approach in which a problem is broken-down into its component functions, which are further divided into sub-functions until the function level suitable for solving the problem is reached.

Other approaches of solving the inverse problem are provided in [30]-[33], and [34]. It is apparent from the literature that there is no unified approach or method to solve the entire inverse problems, as such, what is being presented here is a representation of general methods to the solution for inverse problem.

B. Case-based Reasoning and Inverse Problems

There exist very few publications specifically on Case-Based Reasoning and Inverse Problem in the financial

domain, although the literature has some works done specifically using the Case Base Reasoning methodology and some other combining the methodology with other approaches, for instance, [35] presented a work on CBR and the Inverse Problems to improve usability of numerical models, and to show that CBR models are capable of giving fast answers to questions that are otherwise difficult or impossible to formulate through numerical models. Some selected works involving the use of Inverse problems are presented in [28], [29], [36]-[40]. The IP was reportedly applied in biometric technology by generating synthetic (false) biometric information as a way of training biometric systems and preparing them against attack from false data. Others are seen in [41]-[43]. The combination the CBR with other methodology is reported in [44]. Also, adjacent to the concept proposed in this work is provided in [9] where an early-warning approach for detection of bubbles is proposed. That work uses the minimum-volume ellipsoids clustering method and Radon transform which tackles the bubble concept geometrically by determining and evaluating ellipsoids. Although closely related, work differs from IPCBR both in terms of its approach and concept. The IPCBR model utilizes stock data in its frameworks in CBR and IP phases: the CBR phase uses clustering to retrieve similar bubble structure while the IP phase uses sentiments analysis to assist explanation of bubble occurrences.

IV. PROPOSED FRAMEWORK

Owing the complexity of the problem at hand, we will attempt to tackle the problem by defining and solving its simplified forward problem and then with a clear definition of this, the solution of which will then be an input to the inverse problem. As such, the ensemble is made up two sections: The Case Based Reasoning Model and the Inverse Problem Model. First of all, the CBR model evaluates the potential indicators of all the stocks and output with potentially high yielding stocks with respect to the predefined criteria as a preselected stock set. Secondly, input this stock set, together with its corresponding indicators into the inverse Problem Model. The holistic framework is detailed in Fig. 4.

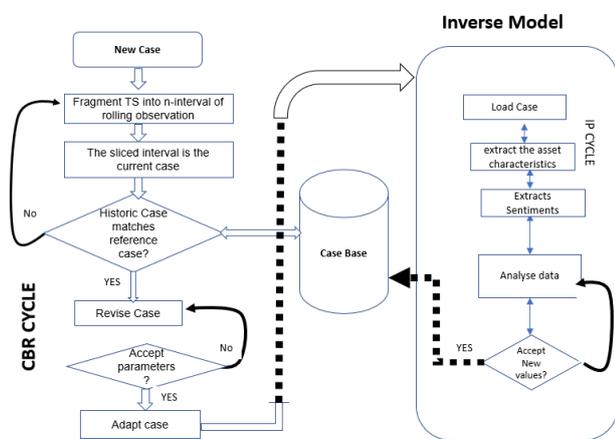


Fig. 4. CBR/IP Model framework.

With all the afore stated assumptions that defines our descriptive model, we aim to arrive at a representative of the descriptive model by calibrating the model parameters of the

seed model through Case Base knowledge, which will be used to initially populate our case base. This involves representing our bubble model in a case structure which is made of historical stock projections represented by a set of points, where each point was given with the time of measuring and the equivalent stock volume. It follows from this that these processes could be represented as curves. Then follows a Pattern Matching phase which entails the process of automatically mapping an input representation for an entity or relationship to an output category. This involves Using the new model perform pattern recognition to identify new instances that fit into the model with the use of appropriate similarity metric. Meaning a new case (the reference case) from the data of new stock which constitutes the input profile is compared with in the case base to retrieve all cases having similar structure and the best-case match is identified. For this investigation, the Dynamic Time Warping will be considered as it is proven to be effective in finding distances Time Series [46] and also because most classic data mining algorithms do not perform or scale well on time series data. If a perfect match is found, then the complete cycle of the CBR will be adopted and solutions adapted, otherwise, a new problem case will be reformulated. The output of this phase signifies the end of the forward problem and the solution then used as a seed for the Inverse Problem phase. IP implementation requires taking the population of newly identified structure which satisfies our bubbles and extract the asset characteristics around the time of the occurrence. Identify correlation between such characteristics and Forward Problem model parameters in order to derive stochastic description of the factors that accompany the above bubbles. The output of this phase will as well be stored in the Knowledge base for easy recommendation.

V. IMPLEMENTATION

In order to evaluate the proposed methodology, several experiments were designed and performed using the proprietary software [45]-[47]. These experiments were ran on benchmark datasets drawn from New York Stock Exchange (NYSE) data obtained from Yahoo finance. For the purpose of this report, we consider the monthly stock prices of sixteen companies. The time series data set ranges from 2000 to 2018. which is made up of 4779 number of observations. The procedure is to collect the daily index historical data and to pre-process them for outliers, missing values, and standardization of the data.

Our focus is on the Adjusted price rather than the Closing price. Although they both provide different information that can be used for analysis, the closing price is the raw price and only indicates end of sales price whereas the Adjusted price mirrors stock value after adjustments for any corporate actions like all applicable splits and dividend distributions. We closely follow the approach presented in [46] creating data by sliding a fixed-length time window from time t_b to t_e .

Resulting in $N = t_e - t_b$ time series created with a specified window length w_{tr} .

$$\begin{aligned}
 s1 &: p1, p2, \dots, p_{w_{tr}} \\
 s2 &: p2, p3, \dots, p_{w_{tr} + 1} \\
 &\dots \\
 sN &: pN, pN + 1, \dots, p_{w_{tr} + N - 1}
 \end{aligned}$$

where $p_i (i = 1, 2, \dots, w_{tr} + N - 1)$ are stock prices at time i . This transformed to creating an N by w_{tr} matrix or a data set with N data records and w_{tr} attributes of continuous values such that data mining methods can be applied directly [47].

A Results and Discussions

We determined the optimal number of components which capture the greatest amount of variance in the data by making a bar plot of the whole dataset with all the components. (Fig. 5).

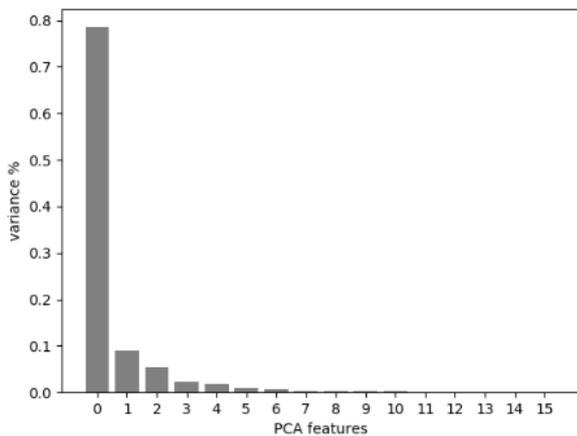


Fig. 5. Variance Bar plot

Figure 6 shows that the first two components explain the majority of the variance in our data. The plot shows a marginal drop after the first component; with this in mind, we plotted just the first two components with the results showing clearly four distinguishable clusters. We can then deduce from this that the stocks can be classified. But we need to visualize the rest of the reduced dataset with much greater granularity, and for this, we applied a combination of Principal Components Analysis (PCA) and K-means Clustering to improve segmentation results [50]. When we visualize the raw data set on the two numerical features TOT and CVX, in figure 5, where the graph represents all points in our current dataset, which our K-means algorithm will aim to segment. Our segmentation model will be based on similarities and dissimilarities between individuals on the features that characterize them.

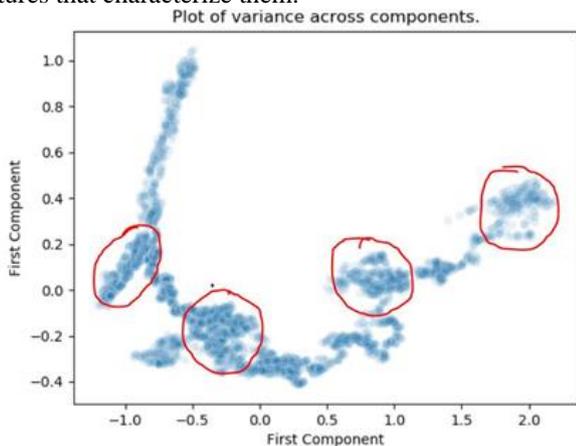


Fig. 6. Plot of variance across components.

Further dimensionality reduction was performed by first fitting our standardized data using PCA and then making a cumulative variance plot to ascertain the number of many features to keep based on the plot, and finally use the selected components to perform PCA (Fig. 7 a). The elbow method was used to guide the specific number of clusters to keep by fitting these principal components to the k-means algorithm. Our experiment tested the algorithm with up to 40 clusters and the results displayed in Figure 5

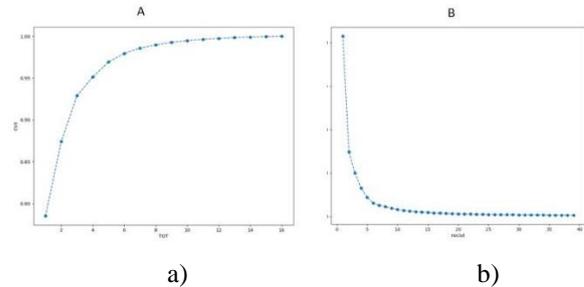


Fig. 7: Plot of cumulative variance variance(a) and elbow point (b)

The elbow plot in Figure 7 b indicates the percentage of variance explained, but in slightly different terms, i.e., as a function of the number of clusters. We can deduce from the graph that after 5 clusters (the elbow point), neither the shift in the value of inertia nor the variance of the rest of the data after is substantial any longer. The partial output from of the algorithms is presented as a data frame in Figure 6.

	TOT	AAPL	ASPEN	BA	...	cmp2	cmp3,	cmp4	p_sc
0	1.000000	1.000000	1.000000	1.000000	...	0.713293	-0.280453	0.121949	0
1	0.970865	0.915690	0.992693	0.998445	...	0.677965	-0.258133	0.096498	0
2	0.970865	0.929090	1.010438	1.060654	...	0.675420	-0.275509	0.104395	0
3	0.949248	0.848688	1.012526	1.071540	...	0.656872	-0.270687	0.071284	0
4	0.946429	0.888889	0.997912	1.102644	...	0.707742	-0.311069	0.106394	0
5	0.936091	0.873255	1.077244	1.087092	...	0.730935	-0.314551	0.111441	0
6	0.943609	0.828588	1.183716	1.066874	...	0.728713	-0.306069	0.087798	0
7	0.928572	0.778894	1.179541	1.071540	...	0.732589	-0.327163	0.083901	0
8	0.921053	0.864322	1.177453	1.054433	...	0.738438	-0.313810	0.084417	0
9	0.934210	0.897264	1.175365	1.094868	...	0.767724	-0.305633	0.087348	0

Fig. 8. PCA and K-means clusters Dataframe.

Figure 9 provides a visualization of our clusters on a 2D plane based on the first two components identified.

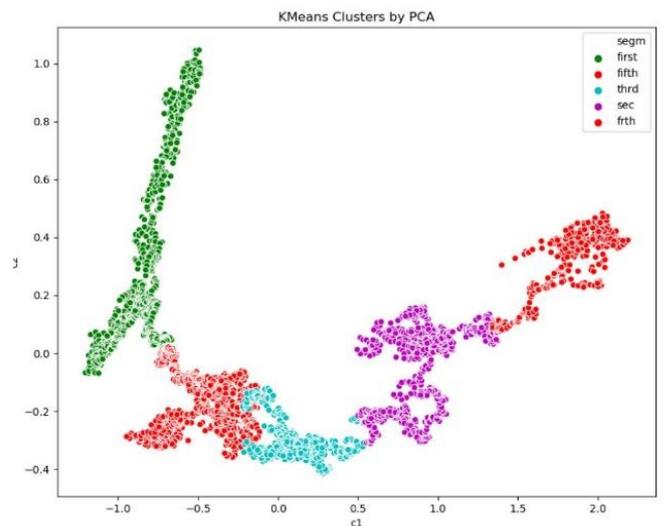


Fig. 9. k-means clusters by PCA.

Although some overlap between the clusters is observed,

this is rather typical, and separate clusters are rather clearly identifiable.

B. Evaluation

Evaluation of unsupervised learning is difficult as there is no goal model to compare with. With K-Means being an unsupervised clustering algorithm where a predicted label does not exist, it is improper to directly apply accuracy score to K-Means clustering evaluation. We therefore decided to use the silhouette coefficient, which is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample [51] to evaluate the performance of our clustering. We then compare the result with a different metric, the Calinski Harabasz [52] Score (also known as Variance Ratio Criterion).

Number of Clusters	Silhouette Coefficient	Calinski Harabasz Score
2	0.5482439	7559.362234
3	0.474617696	6236.356012
4	0.423295284	6614.672999
5	0.430986508	7121.192679
6	0.462071168	7925.66647
7	0.442478772	7788.24215
8	0.399104612	7581.796029
9	0.397991711	7489.326241

Fig. 10. Silhouette coefficient and Calinski Harabasz Score.

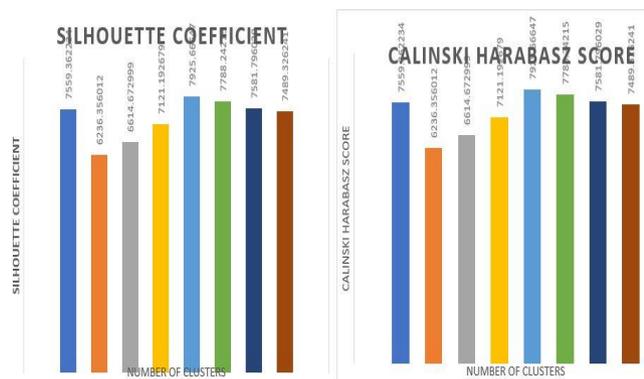


Fig. 11: Plot of silhouette score and Calinski Harabasz against number of clusters.

Both Fig. 10 and 11 shows the performance of the K-means cluster using silhouette coefficient and the Calinski Harabasz Score: the highest score gives the best performance at 2 clusters, but the scatter plots revealed increased overlap between the clusters at this point. Also, we noticed that the scores reduce after this and present a peak at 6 clusters as shown in both metrics, which evidently gives the best performance.

VI. CONCLUSION AND FUTURE WORK

This paper presents the IPCBR approach that uses an AI ensemble of CBR and Inverse Problem formulation, to describe, identify and ultimately predict abnormal fluctuations in stock markets, widely known as bubbles. The proposed framework uses a flexible query engine based on historical time series data and seeks to identify price fluctuations using the Adjusted Close price. The result is able to answer a basic question raised by performing case retrieval, of whether fluctuations in stocks can be classified

based on the Adjusted close features. Decomposing all the features down to principal components and then envisioning the clusters in those principal components using k-means indicates that the stocks can be successfully classified based on this.

For future work, effort is put into formulating the complete forward problem by select representative candidate object which has specified 'bubble' characteristics from the time series dataset based on the objects' degree in their neighbour network through clustering before applying CBR through computing the similarities from characteristics of the case using controlled experiment.

REFERENCES

- [1] Witten, I.H., Frank, E., Hall, M.A.: Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). (2011).
- [2] Huang, S.H.: Supervised feature selection: A tutorial. *Artificial Intelligence Re- search* 4(2) (2015).
- [3] Shokouhi, S.V., Skalle, P., Aamodt, A.: An overview of case-based reasoning applications in drilling engineering. *Artificial Intelligence Review* 41(3) (2014) 317–329.
- [4] Ou, P., Wang, H.: Prediction of stock market index movement by ten data mining techniques. *Modern Applied Science* 3(12) (2009) 28–42
- [5] Milosevic, N.: Equity forecast: Predicting long term stock price movement using machine learning. *Journal of Economics Library* 3(2) (2016) 8.
- [6] Xu, Y., Cohen, S.B.: Stock Movement Prediction from Tweets and Historical Prices. *Acl* (2018) 1–10.
- [7] Barberis, N., Greenwood, R., Jin, L., Shleifer, A.: *Extrapolation and Bubbles* (2017).
- [8] Sornette, D., Cauwels, P.: Financial bubbles: mechanisms and diagnostics. (January) (2014) 1–24.
- [9] Jenny Freeman, T.Y.: Early warning on stock market bubbles via methods of optimization, clustering and inverse problems. *Annals of Operations Research* 260(1-2) (2018) 293–320.
- [10] Jiang, Z.Q., Zhou, W.X., Sornette, D., Woodard, R., Bastiaensen, K., Cauwels, P.: Bubble diagnosis and prediction of the 2005-2007 and 2008-2009 Chinese stock market bubbles. *Journal of Economic Behavior and Organization* 74(3) (2010) 149–162.
- [11] Katja Taipalus: Detecting asset price bubbles with time-series methods Detecting asset price bubbles with time-series methods. (2012).
- [12] Dvhg, D.V.H., Eulg, H., Iru, V., Ixqgdphqwd, R.Q., Whfkqldo, D.Q.G., Vlv, D., Ghflvrlq, I.R.U.: A Case-Based Reasoning-Decision Tree Hybrid System for Stock Selection. 10(6) (2016) 1181–1187.
- [13] Zhou, W.X.: Should Monetary Policy Target Asset Bubbles? –. (2007)
- [14] Ince, H.: Short term stock selection with case-based reasoning technique. *Applied Soft Computing Journal* 22 (2014) 205–212.
- [15] Aamodt, A., Plaza, E.: *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*. AI Communications. IOS Press 7(1) (1994) 39–59.
- [16] Lei, Y., Peng, Y., Ruan, X.: Applying case-based reasoning to cold forging process planning. *Journal of Materials Processing Technology* 112(1) (2001) 12–16.
- [17] Merelli, E., Luck, M.: Technical Forum Group on Agents in Bioinformatics. *Knowledge Engineering Review* 20(2) (2004) 117–125.
- [18] Marketos, G., Padiatitakis, K., Theodoridis, Y., Theodoulidis, B.: Intelligent Stock Market Assistant using Temporal Data Mining. *Citeseer* (May 2014) (1999) 1–11.
- [19] Pecar, B.: Case-based Algorithm for Pattern Recognition and Extrapolation (APRE Method). *SGES/SGAI International Conference on Knowledge Based Systems and Applied Artificial Intelligence* (2002).
- [20] Berthold, M.R., Hoppner, F.: On Clustering Time Series Using Euclidean Distance and Pearson Correlation. (2016).
- [21] Momeni, M., Mohseni, M., Soofi, M.: Clustering Stock Market Companies via K-Means Algorithm. *Kuwait Chapter of Arabian Journal of Business and Management Review* 4(5) (2015) 1–10.
- [22] Datta, T., Ghosh, I.: Using Clustering Method to Understand Indian Stock Market Volatility. *Communications on Applied Electronics* 2(6) (2015) 35–44.
- [23] L'opez, B.: *Case-Based Reasoning: A Concise Introduction*. Synthesis Lectures on Artificial Intelligence and Machine Learning 7(1) (2013) 1–103.

- [24] H., S., Elmogy, M.: Case Based Reasoning: Case Representation Methodologies. *International Journal of Advanced Computer Science and Applications* 6(11) (2015) 192–208.
- [25] Argoul, P.: Overview of Inverse Problems, Parameter Identification in Civil Engineering. (2012) 1–13.
- [26] Bal, G.: Introduction to Inverse Problems. (2012).
- [27] Iulian, M.M.: Direct Problems and Inverse Problems in Biometric Systems. III(5) (2013) 1–14.
- [28] Ulrych, T.J., Sacchi, M.D., Woodbury, A.: A Bayes tour of inversion: A tutorial. *Geophysics* 66(1) (2001) 55–69.
- [29] Yao, Z., Eddy, W.F.: A statistical approach to the inverse problem in magnetoencephalography. *The Annals of Applied Statistics* 8(2) (2014) 1119–1144.
- [30] Jeewana, C.: Domain Decomposition Based Algorithms for Some Inverse Problems. (August) (2000).
- [31] Elshafey, I.M.: Neural network approach for solving inverse problems. (1991).
- [32] Yu, G., Sapiro, G., Mallat, S.: Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* 21(5) (2012) 2481–2499.
- [33] Al-Jamal, M.F.: Numerical solutions of elliptic inverse problems via the equation error method. (2012).
- [34] Knight, B., Petridis, M., Woon, F.L.: Case Selection and Interpolation in CBR Retrieval. 10(1) (2010) 31–38
- [35] Sever, A.: A Machine Learning Algorithm Based on Inverse Problems for Software Requirements Selection. *Journal of Advances in Mathematics and Computer Science* 23(2) (2017) 1–16.
- [36] Michael V, M.V.K.: The Cost Minimizing Inverse Classification Problem : A Genetic Algorithm Approach. *Science Direct Volume* 29,(3) (2000) 283–300.
- [37] Acevedo, N.I.A., Roberty, N.C.: An explicit formulation for the inverse transport problem using only external detectors – Part I : Computational Modelling. *Transport* 29 (2010) 343–358.
- [38] Sever, A.: An inverse problem approach to pattern recognition in industry. *Applied Computing and Informatics* 11(1) (2015) 1–12.
- [39] Search, H., Journals, C., Contact, A., Iopscience, M., Address, I.P.: Inverse problems in Machine Learning : machine learning : an application Interpretation application to activity interpretation. *Theory and Practice* 012085 (2008).
- [40] Floyd, M.W., Esfandiari, B.: An active approach to automatic case generation. McGinty L., Wilson D.C. (eds) *Case-Based Reasoning Research and Development. ICCBR 2009, LNCS 5650* (2009) 150–164.
- [41] Manzoor, J., Asif, S., Masud, M., Khan, M.J.: Automatic Case Generation for Case-Based Reasoning Systems Using Genetic Algorithms. *2012 Third Global Congress on Intelligent Systems* (2012) 311–314.
- [42] Xing, G., Ding, J., Chai, T., Afshar, P., Wang, H.: Hybrid intelligent parameter estimation based on grey case-based reasoning for laminar cooling process. *Engineering Applications of Artificial Intelligence* 25(2) (2012) 418–429.
- [43] Woon, F.L.: Using CBR to Improve the Usability of Numerical Models By. (2005).
- [44] Lin, Jessica, Sheri Williamson, Kirk Borne, D.D.: Chapter 1 Pattern Recognition in Time Series. (2011).
- [45] Van Rossum, G., Drake Jr, F.L.: Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam (1995)
- [46] McKinney, W.: Data structures for statistical computing in python. In van der Walt, S., Millman, J., eds.: *Proceedings of the 9th Python in Science Conference*. (2010) 51 – 56.
- [47] Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculau, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G.: API design for machine learning software: experiences from the scikit-learn project. In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*. (2013) 108–122.
- [48] He, H., Chen, J., Jin, H., Chen, S.: Stock trend analysis and trading strategy. *Proceedings of the 9th Joint Conference on Information Sciences, JCIS 2006* 2006(January) (2006).
- [49] Chen, S.H., Wang, P.P., Kuo, T.W.: Computational intelligence in economics and finance: Volume II. *Computational Intelligence in Economics and Finance: Volume II* (January) (2007) 1–227.
- [50] Ding, C., He, X.: K-means clustering via principal component analysis. *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004* (2004) 225–232.
- [51] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [52] Baarsch, J., Celebi, M.E.: Investigation of internal validity measures for K-means clustering. *Lecture Notes in Engineering and Computer Science* 2195 (2012) 471–476.