



An Ensemble Method: Case-Based Reasoning and the Inverse Problems in Investigating Financial Bubbles

Francis Ekpenyong¹(✉), Georgios Samakovitis², Stelios Kapetanakis¹,
and Miltos Petridis³

¹ University of Brighton, Brighton, UK
{F.Ekpenyong,S.Kapetanakis}@brighton.ac.uk

² University of Greenwich, London, UK
G.Samakovitis@gre.ac.uk

³ Middlesex University, London, UK
M.Petridis@mdx.ac.uk

Abstract. This paper presents an ensemble approach and model; IPCBR, 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) in time series datasets from historical stock market prices. The framework proposes to use a rich set of past observations and geometric pattern description and then applies a CBR to formulate the forward problem; Inverse Problem formulation is then applied to identify a set of parameters that can statistically be associated with the occurrence of the observed patterns.

The technique brings a novel perspective to the problem of asset bubbles predictability. Conventional research practice uses traditional forward approaches to predict abnormal fluctuations in financial time series; conversely, this work proposes a formative strategy aimed to determine the causes of behaviour, rather than predict future time series points. This suggests a deviation from the existing research trend.

Keywords: Case-based reasoning · Inverse problems · Asset bubble · Machine learning · Time series

1 Introduction

Asset value predictability has always been one of the thorny research issues in Finance. While real uses of Artificial Intelligence in the field are as old as the mid-80s, it has been the more topical technological developments that appear to impact direct real time implementation of machine learning in the field. Relevant researches have delivered statistical prediction models [1–3] which, though promising improved predictive capability, they are yet to receive wider acceptance in practice. The reasons for that may stem from the stochastic nature of asset fluctuations in the market which makes it difficult to build reliable models.

A major concern arises with the increase in complexity of the problem, so it becomes very difficult to mathematically formulate the problems, which often leads the choice of parameters to be set by heuristics. This in turn contributes to further 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. The large scale impact of asset price bubble around many historic periods of economic downturn and instability, coupled with the difficulty of identifying a bubble and a general misunderstanding of bubbles, warrants further research study as evident from [4] who reported that econometric detection of asset price bubbles cannot be achieved with a satisfactory degree of certainty despite all the research advancements.

These and other reports demonstrate the need to enhance the explanatory capability of existing artificial intelligence models; to that end, we propose an ensemble IPCBR model that aims to

- (i) use CBR to deliver a more robust representation of asset value fluctuation patterns (and their subsequent classification as asset bubbles) and;
- (ii) implement an Inverse Problem formulation approach to identify the factors that are most likely causes of those patterns.

CBR has been successfully applied to various financial and management domains such as supply chain management and scheduling [5–8], stock selection [9], bond rating [10], bankruptcy prediction and credit analysis [11], and time series prediction [12,13]. The term “inverse problem” which first appeared in the late 1960s, has witnessed a great drift from its original use in geophysics to determine the unknown parameters through input/output or cause-effect experiments, to a contemporary “inverse problems” 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 [14].

Inverse Problem approaches have been successfully applied mainly in science and engineering fields, and provide a truly multidisciplinary platform where related problems from different disciplines can be studied under a common approach with comparable results [15]; these includes Pattern recognition [16], Civil engineering [14] soil hydraulics, [17] computer vision [16], Real-time decision support system [18] machine learning [19,20] and (big) data analysis in general, amongst others. To the small extent, used in financial applications [21] to provide early warning signalling. In spite of the occurrence frequencies of stock market bubbles, the inverse problem approach may contribute to identifying a defining sets of parameters that statistically cause these bubbles.

To tackle the inverse problem, the forward model needs to be first created; to do this, a knowledge mining model is created starting from developing a generic model which covers relevant information on historical stock transactions including the applied results.

The outcome of this will then be used as a case base for standard Case-based Reasoning process, and will be evaluated against a known episodic (real) data and human expert advice. The dataset is drawn from the world largest

recorded stock market repository; New York Stock Exchange (NYSE). About 2,800 companies are listed on the NYSE. In this experiment, we'll look at the daily stock prices of six companies namely: IBM, General Electric (GE), SP500, Tesla, Microsoft and Oracle. The data used in this problem comes from one of the widely used repositories, Yahoo finance.

The remainder of this paper is organised as follows: Sect. 2 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. A simple stochastic asset bubble model is then proposed in Sect. 3 with a brief review of the rational bubble model, which is the theoretical backbone of rational bubble tests, while Sect. 4 expands on a structural geometric representation of the model, which is proposed to act as a base description for our CBR training and subsequent implementation. Section 5 provides an articulation of the overall model to be used where also the Inverse Problem formulation component is discussed. The paper closes with a critical discussion of the major contributions this work intends to deliver, and a set of relevant concluding observations.

2 Prior Research

This section introduces the scope of our research in the well-documented area of stock market bubbles and also examines the general concept of CBR and the IP, with particular reference to previous research. That later provides the grounds for proposing the CBR/IP ensemble approach as a method for identifying the causative parameters in stock market bubbles.

2.1 Asset Bubbles

Various definitions are available in the Finance literature for asset bubbles [22–24], however, broadly described, asset bubbles are significant growths in the market that are not based on substantial change in market or industry performance, and usually escalate and equally dissipate with little or no warning.

Bubbles are often defined relative to the fundamental value of an asset [25, 26]. This can occur if investors hold the asset because they believe that they can sell it at an even higher price to some other investor even though the asset's price exceeds its fundamental value [21]. Detecting a bubble in real time is quite challenging because what attributes to the fundamental value is difficult to pin down. Although every bubble differs in their initiation and specific details, there is a trend in pattern in which informed assumption can be derived this trend makes is possible to recognise bubbles in advance because. Creating a more efficient and effective system that can analyse these fluctuations in patterns can give investors a competitive advantage over others as they can identify stocks with potentials of bubbles with minimum efforts.

A large and growing number of papers propose methods of detecting asset bubbles [21, 23, 27–29].

Many machine learning algorithms have been adopted over the years towards attempting to predict asset bubbles [9, 30]. However, while both academic and trade literature have long been examining their occurrence [23, 26, 31, 32], that extensive literature falls well beyond the scope of this work; for our purposes 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 that occurs in a short period of time, and which is followed by an equally extra-ordinary price decay in a comparably short period’.

The motivation for this relatively constrained focus is evident: due to the extremely convoluted nature of asset bubbles as these have been historically manifested and documented in Finance, Accounting and Economics literature, any attempt to address the phenomenon in its fullness in engineering terms would require a very large set of features and data points and involve infeasible computational complexity.

2.2 Case Based Reasoning

It is assumed that a decision-maker can only learn from experience, by evaluating an act based on its past performance in similar circumstances which informs the application of case-based Reasoning(CBR) Case-based reasoning [33] is an Artificial Intelligence (AI) technique that supports the capability of reasoning and learning in advanced Decision Support System (DSS)[34, 35]. It is a paradigm for combining problem solving and learning which is analogous to problem solving that compares new cases with previous indexes cases. CBR provides two main functions: storage of new cases in the database through indexation module and searching the indexes cases with the similarities of new cases in case retrieval module [36, 37]. The case-based reasoning methodology incorporates four main stages [33, 38]

- Retrieve: given a target problem, retrieve from the case memory, cases that are most relevant and promise to proffer solution to the target case.
- Reuse: the solutions of the best; map the solution from the previous case to the target situation, test the new solution in the real world or perform a simulation, and if necessary.
- Revise: the solution provided by the query case is evaluated and information about whether the solution has or has not provided a desired outcome is gathered.
- Retain: After the solution has been successfully adapted to the target problem, the new problem-solving experience can be stored or not stored in memory, depending on the revise outcomes and the CBR policy regarding case retention.

A CBR cycle is shown in Fig. 1

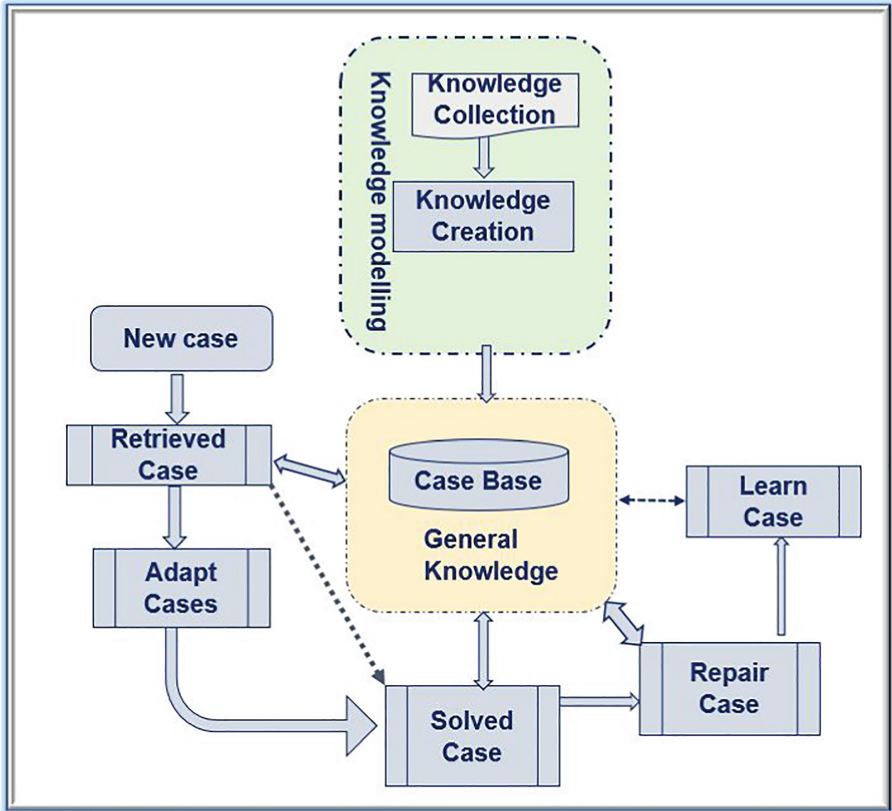


Fig. 1. A modified CBR cycle [33]

2.3 Case Representation

Case representation is a fundamental issue in a Case-based Reasoning methodology. Despite being a plausible and promising data mining methodology, CBR is seldom used in time series domains. This is because the use of case-based reasoning for time series processing introduces some unaccustomed features that do not exist in the processing of the traditional “attribute-value” data representation. Also, direct manipulation of continuous, high dimensional data which involves very long sequences with variable lengths is extremely difficult [39]. In order to make a good case representation, two approaches can be used, namely: the cases represented by succession of points and the cases represented by relations between temporal intervals. Our focus will be on the latter.

2.4 Definition of Cases in the Time Series Context

Case formation tends to be domain specific, and there seems to exist no hard and fast rule in case formation. In our case, we will consider forming a library pattern

of observations and treating every group as a case category. Our representation will mimic the concept proposed in [40]. Where the entire Time series is split into smaller sequences of patterns, each of which is then treated as a case.

This could be achieved by decomposing the series into a sequence of rolling observation patterns or rolling windows. In which case, every observation in the pattern constitutes the case. A case can attain a predefined upward, steady or declining patterns as shown in Fig. 2

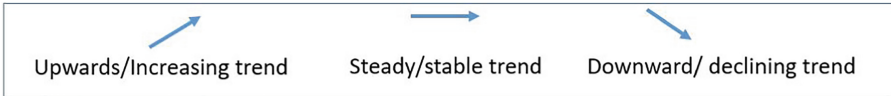


Fig. 2. Sample case patterns

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 appropriate selected algorithm makes it possible to discover a specific relation to the pattern.

2.5 Computing Similarities in Time Series

Similarity measure is the most essential ingredient of time series clustering and classification systems, and one of the first decisions to be made in order to establish how the distance between two independent vectors must be measured [41]. Because of this importance, numerous approaches to estimate time series similarity have been proposed. Among these, Longest common subsequence (LCS) [42, 43] Histogram-based similarity measure [44] Cubic Spline [45], dynamic Time Wrapping [46, 47] have been extensively used.

Similarity in real sense is subjective, highly dependent on the domain and application. It is often measured in the range 0 to 1 [0,1], where 1 indicates the maximum of similarity.

Similarity between two numbers x and y can be represented as:

$$nSim(X, Y) = 1 - \frac{|x - y|}{|x| + |y|} \tag{1}$$

When considering two time series $X = x_1, \dots, x_n, Y = y_1, \dots, y_n$, some similarity measures that could be used are:

mean similarity defined as:

$$MSim(X, Y) = \frac{1}{n} \sum_{i=1}^n nSim(x_i, y_i) \tag{2}$$

Root mean Square similarity defined as:

$$RMSim(X, Y) = \sqrt{\frac{1}{n} \sum_{i=1}^n Sim(x_i, y_i)^2} \quad (3)$$

Peak similarity defined as:

$$PSim(X, Y) = \frac{1}{n} \sum_{i=1}^n \left[\frac{|x_i - y_i|}{2max(|x_i|, |y_i|)} \right] \quad (4)$$

Despite the sporadic introduction of various methods of measuring similarity, the challenge of determining the best method for assignment of attributes' weight value in CBR still needs to be addressed [37].

Euclidean distance is by far the most popular distance measure in data mining, and it has the advantage of being a distance metric. However, a major demerit of Euclidean distance is that it requires that both input sequences be of the same length, and it is sensitive to distortions, e.g. shifting, along the time axis. Such a problem can generally be handled by elastic distance measures such as Dynamic Time Warping (DTW)

In this research, the dynamic time warping (DTW) distance was chosen owing to the fact that it overcomes some limitations of other distance metrics by using dynamic programming technique to determine the best alignment that will produce the optimal distance. It is an extensively used technique in speech recognition [48, 49] and many other domains, including Time Series analysis [46, 47].

2.6 Inverse Problem

Inverse problems (also called model inversion) arise in many fields, where one tries to find a model that typically approximates observational data. Any inverse theory requirement is to relate a physical parameter “u” that describes a model to acquire observations making up some set of data “f”. Assuming there is a clear picture of the underlying concept of the model, then an operator can be assigned a relation or mapping u to f through the equation:

$$f = Ku \quad (5)$$

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. It can be referred to as the Green's function because of the analogy with the continuous function case:

$$f(dx) = K(x, t)u(t)dx \quad (6)$$

In a case where the experiment is drawn from i observations and k model parameters will be:

$$\text{data: } f = [f_1, f_2, \dots, f_i]^T$$

and model parameters: $u = [u_1, u_2, \dots, u_j]^T$

with d and m representing I and J dimensional column vectors, respectively, and T denoting the transpose.

Most relevant application find it very difficult to invert the forward problem for some obvious reasons; either a (unique) inverse model simply does not exist, or a small perturbation of the system causes a relatively large change in the exact solution. In the sense of Hadamard the problem above is called well-posed under the conditions of:

- Existence: \forall input data there exists a solution of the problem, i.e. for all $f \in$ there exists a $u \in U$ with $Ku = f$.
- Uniqueness: \forall input data, this existing solution is unique meaning $u \neq v$ implies $Kv \neq f$
- Stability: the solution of the problem depends continuously on the input datum, i.e. $\forall Uk_{K \in N}$ with $Ku_k \mapsto f$ we have $u_k \mapsto u$.

The well-posedness of a model highly depends on the stated conditions, Violation of any of the conditions results in ill posedness, or approximately ill-posed [50]. One way of finding the inverse of this in the use of Convolution [51], which is widely significant as a physical concept and offers an advantageous starting point for many theoretical developments.

A convolution operation describes the action of an observing instrument when it takes a weighted mean of a physical quantity over a narrow range of a variable. It is widely used in time series analysis as well to represent physical processes.

The convolution of two functions $f(x)$ and $g(x)$ represented as $f(x)*g(x)$ is

$$\int_{-\infty}^{\infty} f(u)g(x - u)du \tag{7}$$

As such a more logical step is to take the forward problem in Eq. 5 and invert it for an estimate of the model parameters f^{est} as

$$f^{est} = K^{inverse}u$$

by performing a Deconvolution to it which could be represented as

$$f = K^{-1}u \tag{8}$$

Alternatively, the equation can be reformulate the problem as

$$f^T K f = K^T u \tag{9}$$

and find the solution as

$$f = [K^T K]^{-1}[K^T u] \tag{10}$$

Considering the large number of reported asset bubbles in stock markets, there appears to exist a rich body of pattern occurrences to allow for applying the Inverse Problem approach to identify a defining set of parameters that statistically cause them.

3 Asset Bubble from Price Theory

This section gives a brief review of the rational bubble model, which is the theoretical backbone of rational bubble tests. A simple linear asset pricing model is employed that draws its arguments from the basic financial theory [26,52], which expresses asset price as a discount factor multiplied by the flows of all future payments relating to the asset [23].

Let P_t denote the price of an asset at time t , and return rate R_t at time t based on the work of [28]. The return rate of the asset in the next period is denoted with R_{Δ} , and its equivalent change in price at a time t_{Δ} to be denoted with P_{Δ} . Then The return rate R_{Δ} of a stock can be expressed as the as the sum of the price change ($P_{\Delta} - P_t$) and the dividend D_{Δ} , adjusted to the price of a stock in period t , given by

$$R_{\Delta} = \frac{P_{\Delta} - P_t + D_{\Delta}}{P_t} \quad (11)$$

The change in price and the dividends becomes apparent only in the period t_{Δ} as they realise, one can take mathematical expectation of Eq. 11 based on available information on period t , this being

$$E_t(R_{\Delta}) = \frac{E(P_{\Delta} - P_t + D_{\Delta}) - P_t}{P_t} = R \quad (12)$$

Where

$E_t(P_t) - P_t + D_{\Delta}) = RP_t$ Rearranging 12 results in

$$P_t = E_t\left(\frac{D_{\Delta}}{1+R}\right) + E_t\left(\frac{P_{\Delta}}{1+R}\right) \quad (13)$$

For periods more than singular denoted by k the forward solution could further be stated as

$$P_t = E_t\left[\sum_{i=1}^k \left(\frac{1}{1+R}\right)^i D_{\Delta}\right] + E_t\left[\sum_{i=1}^k \left(\frac{1}{1+R}\right)^j P_{\Delta}\right] \quad (14)$$

And again, applying the elements of uniqueness to solution for this equation, it is also assumed that the expected discounted value of a stock converges to zero in Eq. 14 under assumption on indefinite amount future periods [28].

$$\text{Lim}_{k \rightarrow \infty} E_t\left[\left(\frac{1}{1+R_{t+k}}\right)^i P_{t+k}\right] = 0 \quad (15)$$

Hence, reducing the forward solution of the stocks fundamental price

$$(P_t^f) = E_t\left[\sum_{i=1}^{\infty} \left(\frac{1}{1+R_{t+i}}\right)^i D_{t+i}\right] \quad (16)$$

where (P_t^f) is the expected discounted value of future dividends. Failure of which would result in infinite number of solutions that can be represented as

$$P_t = (P_t^f) + B_t, \quad (17)$$

where $B_t = E_t \left[\frac{B_{t+\Delta}}{1+R_{t+\Delta}} \right]$

B_t in equation above would present the ‘rational bubble’ as this components value would consist of the expected path of stock price returns.

To promote the financial stability, an effective warning mechanism is always desirable to signal the formation of asset price misalignments. This research provides a methods to accomplish this task by the use of ensemble method, CBR/IP.

The study from [21] presented an early-warning signalling approach for financial bubbles by benefiting from the theory of optimization, of inverse problems and clustering method. The research reported a method which approaches the bubble concept geometrically by determining and evaluating ellipsoids and reported that when the bubble-burst time approaches, the volumes of the ellipsoids gradually decrease and, correspondingly, the figures obtained by Radon transform become more “brilliant” presenting more strongly warning.

The authors of [28] stated that although every bubble has its own unique features, there are some common early symptoms. He further showed that the conventional unit root tests in modified forms can be used to construct early warning indicators for bubbles in financial markets.

4 Structural Representation of the Model

Despite the fact that there are many opinions about bubbles in various literature, one thing is obvious, none of the authors seems to disagree about the theoretical determination of the fundamental asset price. An asset price bubble according to [32] is defined as the difference between two components: the observed market price of a given financial asset, which represents the amount that the marginal buyer is willing to pay, and the asset’s intrinsic or fundamental value, which is defined as the expected sum of future discounted dividends. In trying to give meaning to what a bubble is, let us define what a fundamental value of an asset is. The representation is adopted from the concept given in [52] which starts with a case of an asset that yields a known and fixed stream of dividends. In which case, d_t denote the dividend income paid out by the asset at date t, where t runs from 0 to infinity, and q_t denote the current price of a bond that pays one dollar at date t. Its states the value any trader attaches to the dividend stream from this asset is given by

$$F_t = \sum_{i=0}^{\infty} q_t d_t \quad (18)$$

here, F denotes the fundamental value of a stock. A stock bubble deduces a stock whose price P is not equal to its fundamental value, meaning $P \neq F$. a

bubble case would assume the market price to sell above its fundamental value, in which case $P > F$. Also, on the assumption dividends are uncertain, given a state of the world represented as states in a set Φ which denotes a set of all possible outcomes at a date t , and given that $\tau_t \in \Phi$ refers to a particular state of the world at date t which all dealers hope will occur with a probability say $Prob(\tau_t)$ which determines the value of a dividend at date t given by the fundamental values dealers allocates to the asset in this case is expectation

$$F_t = [\sum_{i=0}^{\infty} q(\tau_t)d(\tau_t)] \tag{19}$$

$$= [\sum_{i=0}^{\infty} \sum_{\tau_t \in \Phi} q(\tau_t)d(\tau_t)] \tag{20}$$

In this case, an asset would be considered a bubble if its price $P > F$, as defined in the equation. The $P <> F$ refers to an asset being “unfairly priced” in the sense of perhaps being valued at discount.

The equation above could be related to a descriptive bubble case adopted from the work of [53], which represents a growing asset prize with respect to time t , shown in Fig. 3. Time here is considered continuous and infinite with periods $t \subset R$.

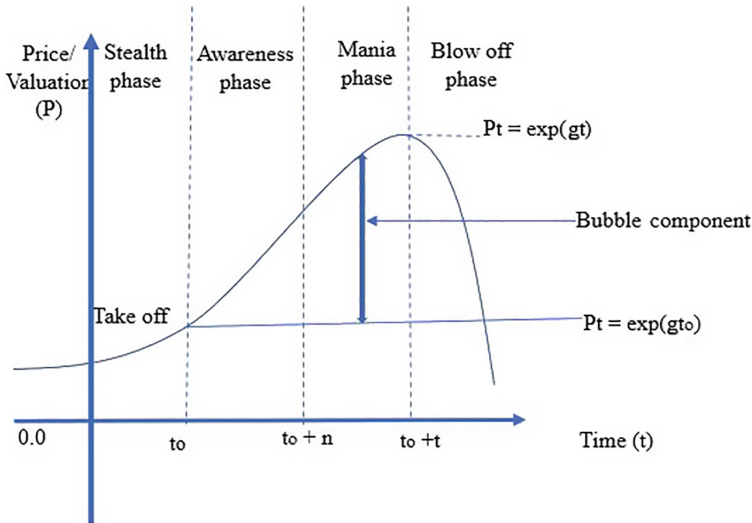


Fig. 3. A smoothed approximation representation of a bubble

The figure shows an initial constant steady growth in asset price based on the fundamental value at some random time t , At a point say t_0 , (take off point or the Stealth phase) the price driven by bubble grows in time value with expectation

(gt_o) , From t_o the asset price pt grows exponentially at $g > 0$, denoting evolving price with a growing expectation given by $p_t = exp(gt)$. Hence the bubble component is denoted by $exp(gt) - exp(go)$ where $t > t_o$. The assumption is that the starting point of a bubble is t_o is discrete as $t_o = 0, \delta, 2\delta, 3\delta, \dots$, where $\delta > 0$ and that t_o is exponentially distributed on $[0, \infty)$ with cumulative distribution $\psi(t_o) = 1 - exp(-\beta t_o)$ [54]. Investors considered are risk neutral investors that have a discount rate of zero, whereby, as long as they hold the assets, they have two choices; either to sell or retain the assets. But when $\alpha \in (0, 1)$ of investors sell their assets, the bubble bursts and the asset price drops to the true value. If fewer than α of the investors sell their assets at time ρ after t_o , the bubble bursts automatically at $t_o + \rho$ but if she sells his assets at t i.e before the bubble bursts, he receives the price in the selling period otherwise he only receives true value $exp(gt_o)$ below the price at $t > t_o$.

5 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.

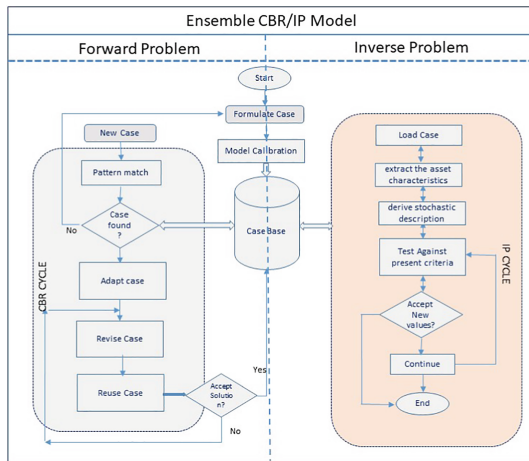


Fig. 4. CBR/IP Model framework

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.

For this investigation, the Dynamic Time Warping will be considered as it is proven to be effective in finding distances Time Series [42] and also because most classic data mining algorithms do not perform or scale well on time series data [55] 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 newly identified structure from the retrieved case and extract the asset characteristics around the time of the occurrence. It then requires that we identify any correlation between such characteristics and the forward problem model parameters in order to derive stochastic description of the factors that accompany the said bubbles. The output of this phase will as well be stored in the Knowledge base for easy recommendation.

6 Conclusion and Future Work

This paper proposes an 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 in temporal constraints.

The ensemble is aimed to 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. The neighbour network is built based on the similarity of time series objects which is measured by suitable similarity metrics.

Once the candidates are chosen, further investigation will be performed to extract the asset characteristics around the time of the occurrence 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.

By capturing such experiences in a new ensemble model, investors can learn lessons about actual challenges to trading assumptions, adequate project preparedness and planned execution and be able to leverage that knowledge for efficient and effective management of future similar transactions.

For future work, the plan will be to create a knowledge pool of distinct types of stock patterns and apply CBR in computing the similarities and characteristics of the case using controlled experiment.

References

1. Ou, P., Wang, H.: Prediction of stock market index movement by ten data mining techniques. *Mod. Appl. Sci.* **3**(12), 28–42 (2009)
2. Milosevic, N.: Equity forecast: predicting long term stock price movement using machine learning. *J. Econ. Libr.* **3**(2), 8 (2016)
3. Xu, Y., Cohen, S.B.: Stock movement prediction from tweets and historical prices. In: *ACL*, pp. 1–10 (2018)
4. Gurkaynak, R.S.: *Econometric Tests of Asset Price Bubbles: Taking Stock Econometric Tests of Asset Price Bubbles* (2005)
5. Dalal, S., Athavale, V.: Analysing supply chain strategy using case-based reasoning. *J. Supply Chain Manage.* **1**, 40 (2012)
6. Kaur, M.: *Inventory Cost Optimization in Supply Chain System Through Case-based Reasoning*, vol. I, no. V, pp. 20–23 (2012)
7. Fu, J., Fu, Y.: Case-based reasoning and multi-agents for cost collaborative management in supply chain. *Procedia Eng.* **29**, 1088–1098 (2012)
8. Lim, S.H.: Case-based reasoning system for prediction of collaboration level using balanced scorecard: a dyadic approach form distributing and manufacturing companies. *J. Comput. Sci.* **6**(9), 9–12 (2006)
9. Ince, H.: Short term stock selection with case-based reasoning technique. *Appl. Soft Comput. J.* **22**, 205–212 (2014)
10. Shin, K.S., Han, I.: A case-based approach using inductive indexing for corporate bond rating. *Decis. Support Syst.* **32**(1), 41–52 (2001)
11. Bryant, S.M.: A case-based reasoning approach to bankruptcy prediction modeling. *Intell. Syst. Account. Finance Manage.* **6**(3), 195–214 (1997)
12. Kurbalija, V., Budimac, Z.: Case-based reasoning framework for generating decision support systems. *Novi Sad J. Math.* **38**(3), 219–226 (2008)
13. Elsayed, A., Hijazi, M.H.A., Coenen, F., Garcia-Finana, M., Sluming, V., Zheng, Y.: Time series case based reasoning for image categorisation. In: *Case-Based Reasoning Research and Development*, pp. 423–436 (2011)
14. Argoul, P.: Overview of Inverse Problems, Parameter Identification in Civil Engineering, pp. 1–13 (2012)
15. Gomez-ramirez, J.: Inverse thinking in economic theory: a radical approach to economic thinking. *Four problems in classical economic modeling* (2003)
16. Sever, A.: An inverse problem approach to pattern recognition in industry. *Appl. Comput. Inf.* **11**(1), 1–12 (2015)
17. Ritter, A., Hupet, F., Mun, R., Lambot, S., Vanclooster, M.: Using inverse methods for estimating soil hydraulic properties from \otimes ; eld data as an alternative to direct methods. *Agric. Water Manag.* **59**, 77–96 (2003)

18. Gundersen, O.E., Srmo, F., Aamot, A., Skalle, P.: A real-time decision support system for high cost oil-well drilling operations. In: Proceedings of the Twenty-Fourth Innovative Applications of Artificial Intelligence Conference A, pp. 2209–2216 (2012)
19. Sever, A.: A machine learning algorithm based on inverse problems for software requirements selection. *J. Adv. Math. Comput. Sci.* **23**(2), 1–16 (2017)
20. Search, H., Journals, C., Contact, A., Iopscience, M., Address, I.P.: Inverse problems Problems in in machine learning: machine learning: an application Interpretation application to activity interpretation. *Theory Pract.* **135**, 012085 (2008)
21. Kürüm, E., Weber, G.W., Iyigun, C.: Early warning on stock market bubbles via methods of optimization, clustering and inverse problems. *Ann. Oper. Res.* **260**(1–2), 293–320 (2018)
22. Herzog, B.: An econophysics model of financial bubbles. *Nat. Sci.* **7**(7), 55–63 (2007)
23. Kubicová, I., Komárek, L.: The classification and identification. *Finance a úvěr-Czech J. Econ. Finan.* **61**, 1(403), 34–48 (2011)
24. Martin, A., Ventura, J.: Economic growth with bubbles. *Am. Econ. Rev.* **102**(6), 3033–3058 (2012)
25. Barberis, N., Greenwood, R., Jin, L., Shleifer, A.: Extrapolation and Bubbles (2017)
26. Sornette, D., Cauwels, P.: Financial Bubbles: Mechanisms and Diagnostics, pp. 1–24, January 2014
27. 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. *J. Econ. Behav. Organ.* **74**(3), 149–162 (2010)
28. Taipalus, K.: Detecting Asset Price Bubbles with Time-series Methods (2012)
29. Dvhg, D.V.H., et al.: A case-based reasoning-decision tree hybrid system for stock selection. *Int. J. Comput. Inf. Eng.* **10**(6), 1181–1187 (2016)
30. Zhou, W.X.: Should Monetary Policy Target Asset Bubbles? (2007)
31. Press, P., Profit, T.: Bursting Bubbles: Finance, Crisis and the Efficient Market Hypothesis. *The Profit Doctrine*, pp. 125–146 (2017)
32. Nedelcu, S.: Mathematical models for financial bubbles. Ph.D thesis (2014)
33. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Commun.* **7**(1), 39–59 (1994)
34. Kolodner, J.L.: Case-based reasoning. In: *The Cambridge Handbook Of: The Learning Sciences*, pp. 225–242 (2006)
35. López, B.: Case-based reasoning: a concise introduction. *Synth. Lect. Artif. Intell. Mach. Learn.* **7**(1), 1–103 (2013)
36. Cunningham, P.: A taxonomy of similarity mechanisms for case-based reasoning. *IEEE Trans. Knowl. Data Eng.* **21**(11), 1532–1543 (2009)
37. Ji, S., Park, M., Lee, H., Yoon, Y.: Similarity measurement method of case-based reasoning for conceptual cost estimation. In: Proceedings of the International Conference on Computing in Civil and Building Engineering (2010)
38. El-Sappagh, S.H., Elmogy, M.: Case based reasoning: case representation methodologies. *Int. J. Adv. Comput. Sci. Appl.* **6**(11), 192–208 (2015)
39. Marketos, G., PEDIADITAKIS, K., THEODORIDIS, Y., THEODOULIDIS, B.: Intelligent Stock Market Assistant using Temporal Data Mining. *CiteSeer* (May 2014), pp. 1–11 (1999)
40. Pecar, B.: Case-based algorithm for pattern recognition and extrapolation (APRE Method). In: *SGES/SGAI International Conference on Knowledge Based Systems and Applied Artificial Intelligence* (2002)

41. Iglesias, F., Kastner, W.: Analysis of similarity measures in times series clustering for the discovery of building energy patterns. *Energies* **6**(2), 579–597 (2013)
42. Sengupta, S., Ojha, P., Wang, H., Blackburn, W.: Effectiveness of similarity measures in classification of time series data with intrinsic and extrinsic variability. In: Proceedings of the 11th IEEE International Conference on Cybernetic Intelligent Systems 2012, CIS 2012, pp. 166–171 (2012)
43. Khan, R., Ahmad, M., Zakarya, M.: Longest common subsequence based algorithm for measuring similarity between time series: a new approach. *World Appl. Sci. J.* **24**(9), 1192–1198 (2013)
44. Lin, J., Li, Y.: Finding structural similarity in time series data using bag-of-patterns representation. In: Winslett, M. (ed.) SSDBM 2009. LNCS, vol. 5566, pp. 461–477. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02279-1_33
45. Wongsai, N., Wongsai, S., Huete, A.R.: Annual seasonality extraction using the cubic spline function and decadal trend in temporal daytime MODIS LST data. *Remote Sens.* **9**(12), 1254 (2017)
46. Zhang, X., Liu, J., Du, Y., Lv, T.: A novel clustering method on time series data. *Expert Syst. Appl.* **38**(9), 11891–11900 (2011)
47. Phan, T.t.h., et al.: Dynamic time warping-based imputation for univariate time series data To cite this version: HAL Id : hal-01609256. *Pattern Recogn. Lett.* (2017)
48. Cassidy, S.: *Speech Recognition: Dynamic Time Warping*, vol. 11, p. 2. Department of Computing, Macquarie University (2002)
49. Xihao, S., Miyanaga, Y.: Dynamic time warping for speech recognition with training part to reduce the computation. In: ISSCS 2013 - International Symposium on Signals, Circuits and Systems (2013)
50. Levitan, B.M., Sargsjan, I.S.: *Inverse Problems. Sturm-Liouville and Dirac Operators*, pp. 139–182 (2012)
51. Tarantola, A.: Chapter 1: Introduction 1.1 Inverse theory: what it is and what it does, vol. 1, pp. 1–11. Elsevier Scientific Publishing Company (1987)
52. Barlevy, G.: *Economic Theory and Asset Bubbles*, pp. 44–59 (2007)
53. Asako, Y., Funaki, Y., Ueda, K., Uto, N.: *Centre for Applied Macroeconomic Analysis Symmetric Information Bubbles: Experimental Evidence* (2017)
54. Abreu, D., Brunnermeier, M.K.: Bubbles and crashes. *Econometrica* **71**(1), 173–204 (2003)
55. Lin, J., Williamson, S., Borne, K., DeBarr, D.: Pattern recognition in time series. *Adv. Mach. Learn. Data Min. Astron.* **1**, 617–645 (2012). <https://doi.org/10.1201/b11822-36>