Available online at www.sciencedirect.com

ScienceDirect

Expert Systems with Applications 34 (2008) 2232–2240

Expert Systems with Applications

www.elsevier.com/locate/eswa

Review

Expert systems and evolutionary computing for financial investing: A review

Roy Rada

Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD 21250, United States

Abstract

This innovative, experimental approach to a literature review begins with queries for finance-related articles to the *Expert Systems with Applications* literature database. A classification language is constructed, and retrieved articles are systematic indexed. Hypotheses are offered about the patterns in the distribution of indexed concepts. Results include that authors tended to use expert systems tools in the early 1990s but evolutionary computation tools in mid-2000s. However, the most common financial application area in both the earlier and later periods was financial accounting. One trend in the latest work is to merge both knowledge-based and evolutionary approaches. Opportunities exist to reuse the knowledge bases built for the financial accounting work for speculative investing.

Keywords: Literature review; Finance; Expert systems; Evolutionary computing; Classification language

Contents

E-mail address: rada@umbc.edu

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2007.05.012

1. Introduction

This paper is a literature review of certain papers published in the journal *Expert Systems with Applications* (hereinafter referred to as *ESwA*). That journal was chosen because of its strength in financial applications. The publisher of *ESwA* solicits manuscripts for that journal that address accounting, economics, finance, and stock trading, as well as other topics (Elsevier, 2006a). By contrast, some artificial intelligence journals favor different application areas. For instance, the publisher of the *IEEE Transactions on Evolutionary Computation* declares in its scope for that journal "with emphasis given to engineering systems and scientific applications" and does not explicitly mention accounting, economics, finance, or stock trading (IEEE, 2006).

An expert system is defined as "A program that uses available information, heuristics, and inference to suggest solutions to problems in a particular discipline" (American Heritage Dictionary of the English Language, Fourth Edition, 2000). In some quarters, the concept 'expert system' refers only to knowledge-based systems. However, in other quarters, the concept also represents 'neural networks'. For instance, a popular management information systems textbook says (Oz, 2006): "Rather than containing a set of IF–THEN rules, more sophisticated Expert Systems use programs called neural networks, which are designed to mimic the way a human brain learns". *ESwA*'s web site says that technology applications of interest include not only expert systems technology but also neural networks, genetic algorithms, and others. Neural networks and genetic algorithms are both studied in a discipline called 'evolutionary computation'. This review considers all technologies addressed by papers in *ESwA*.

A search on the *ESwA* web site for articles containing the stem 'survey' or 'review' returns 135 articles. Of these 135 articles, seven are relevant to the notion of a literature review or survey, whereas others used one of those two terms in some other sense, such as a survey being taken of users' reactions to an expert system. Those seven reviews can be summarized as follows:

- O'Leary (1991) reviewed the literature on accounting and noted the need for connections between the inference engine and databases. Fifteen years later, Ezziane (2006) reviewed the literature on expert systems in bioinformatics and, like O'Leary, focused on the need to connect to databases.
- Liao (2005) retrieved articles published between 1995 and 2004 from multiple academic journals. Liao concluded that intelligent technologies have been applied to many problem domains and that the future of the field is in the continual adaptation of the technology to particular domains. In an earlier review by the same author, similar conclusions are reached (Liao, 2003).
- Nedovic and Devedzic (2002) studied a few expert systems in finance and concluded that each system is tailor-made for that application. An earlier paper by one of these authors (Devedzic, 1999) concluded that object-oriented models have become prevalent.
- Vellido, Lisboa, and Vaughan (1999) reviewed neural network applications in accounting, management, and marketing but explicitly excluded from consideration articles addressing financial market applications, such as predicting stock prices, because "an enormous research effort has already been put to tackle these problems . . . needs to be surveyed on its own".

The preceding reviews and observations raise, at least, three questions: What changes have occurred over time in the tools used for financial applications? What financial applications get how much attention? Are knowledge-based and evolution-based methods being integrated?

In particular, the hypotheses about the literature are:

- Expert system technologies were most used in the early 1990s, but evolutionary computation technologies were most used in the early 2000s,
- Financial investing is more often the application area than is financial accounting, and
- Evolutionary computation and expert system techniques will be increasingly combined.

The tests of these hypotheses will provide information upon which a model for research on intelligent, financial systems can be built.

2. Method

This paper's approach to literature review is innovative. This method section will explain how:

- Queries were systematically developed and applied to literature databases,
- A classification language was developed, and
- Retrieved articles were indexed into the terms of the classification language.

This method allows for a literature review to be repeated or extended in the same way that an experiment might be.

2.1. Queries

The following query was posed on the Elsevier Science Direct™ web site (Elsevier, 2006a):

> ISSN(0957-4174) and TITLE-ABSTR-KEY(financ! OR investm!).

The ISSN for *ESwA* is 0957-4174. The exclamation mark "!" represents the wild card function in this retrieval system. To this query, the retrieval system returned 98 documents on September 26, 2006. The accessible database of *ESwA* publications included 190 that were 'to be published' but had not been published on September 26, 2006.

Scopus™ (Elsevier, 2006b) is an Elsevier literature retrieval system that differs in, at least, three ways from the Science Direct system. For one, queries to the Elsevier Science Direct web site do not work against as much metadata as queries addressed to Scopus. For another, Scopus does not include citations that are 'to be published'. Finally, the Scopus wild card symbol is '*', whereas the Science Direct wild card symbol is '!'. The following query was posed to Scopus:

ISSN(0957-4174) AND (ALL(financ*) OR ALL(investm*)).

This query returned 232 publications in September 2006. The Scopus results were merged with the Science Direct results to get 254 unique references (76 were duplicates).

Since this study is looking at trends, the set of 254 references was reduced to those references published:

- between 1991 and 1996 or
- in the first nine months of 2006 or 'to-be-published'.

Those two sets contained, respectively, 40 and 61 publications. Thus, 101 publications from *ESwA* were identified for further study.

2.2. Classification language

In reviewing the journal literature, one might categorize articles. A classification language can supply concepts with which to classify the articles. The terminology that will be used in this article to discuss classification languages includes 'term', 'concept', and 'hierarchy'. A 'term' is a word or group of words having a particular meaning, such as the term 'classification language'. A concept is an idea. One concept may be represented by several terms that are synonymous. For instance, the terms 'indexing language', 'thesaurus', and 'classification language' could be synonymous. A hierarchy is an organization with one or more things at the top and with several things below each other thing and may be depicted as an inverted tree. The concepts (represented as terms) in the classification language are hierarchically organized.

The classification language for this paper will have two top-level concepts called 'Intelligent Tools' and 'Finance'. No pre-existing classification language adequately covers these two top-level concepts. However, some classification languages have components that might be reused.

Dobrev, Strupchanska, and Toutanova (2001) provide a classification of financial instruments and financial organizations. Finance books provide in their Table of Contents a kind of classification. Bodie and Merton's (1998) introductory finance textbook emphasizes the three pillars of finance: the time value of money, asset valuation, and risk management. Books on intelligent financial investing software, such as Bauer (1994) and Kingdon (1997), specify this step-by-step sequence:

- Identify data to use,
- Predict future values of assets, and
- Maintain a portfolio.

From these and other sources, an initial classification language was constructed that included as the children of 'Finance' the concepts 'Data Identification', 'Asset Valuation', and 'Risk Management'.

For the 'Intelligent Tools' sub-tree of the classification language, one source is the *ACM Computing Classification System* (Coulter et al., 1998). Waltz (1989) built an artificial intelligence classification from the *ACM Computing Classification System*. The initial version of this paper's classification language borrowed the artificial intelligence, theory of computation, statistics, pattern recognition, and numerical analysis parts from the *ACM Computing Classification System*.

Two methods to build a classification language for a body of literature are top-down and bottom-up. (Jagerman, 2006); in the

- top-down method a person starts with a conceptual model of the domain and augments that model with parts of existing classification languages (Mili & Rada, 1988) and
- bottom-up method a person indexes articles in the domain and modifies the classification language based on what the articles require. (Rada, Mili, Letourneau, & Johnston, 1988).

This author began with the top-down method and followed with the bottom-up method.

2.3. Indexing

Through his online subscription to Science Direct™, the author obtained digital copies of the 101 articles corresponding to the 101 retrieved citations. He then read the 101 articles to determine which were relevant to finance. Some articles did not address finance, such as one article about customer relationship management, another about breast cancer diagnosis, and another about semiconductor manufacturing; these were removed from further consideration. Finally, a subset of articles about 'financial accounting' was removed for a subtle reason. An article that considered 'financial accounting' relative to an individual was considered not relevant to this study. For instance, an article was considered not relevant when it addressed the credit rating of individuals. The 'financial accounting' addressed in the relevant articles was for the decision to be made about an organization. For instance, an article about how a bank might determine whether a company that applies for a loan is loan-worthy was considered a relevant article.

Thirty-six articles were deemed irrelevant. That left 65 articles in the relevant set. The 65 relevant citations from *ESwA* were exported from Science Direct™ into EndNote

(EndNote, 2006). These citations included fields for author, title, date, volume, issue, pages, and abstract. Two fields were added to EndNote to store:

- A 'Finance' concept and
- An 'Intelligent Tool' concept.

For each citation, exactly one concept from the classification language 'Finance' tree was inserted in that citation's 'Finance' field in EndNote. Similarly, a concept from the classification language 'Intelligent Tool' tree was inserted into each citation's 'Intelligent Tool' field.

Many articles addressed more than one 'Finance' concept and more than one 'Intelligent Tools' concept. However, the intention in indexing was to capture the one most prominent 'Finance' concept and the one most prominent 'Intelligent Tools' topic. For instance, if an article primarily described a 'neural network' but secondarily described a 'genetic algorithm', then the article's 'Intelligent Tools' field would be assigned the concept 'neural network' and not the concept 'genetic algorithm'. When an article equally addressed multiple 'Intelligent Tools' concepts, then the lowest, common ancestor of those concepts from the classification language was placed into the 'Intelligent Tool' field in the EndNote citation. For instance, if the article equally emphasized a genetic algorithm and a neural network, then the concept used to index the article was the parent of 'neural network' and 'genetic algorithm', namely 'evolutionary computing'. Admittedly, this indexing was subjective, and no other indexers were employed to test for consistency in indexing.

3. Results

The classification language is one result of this research and is described further in the next subsection. The results of the indexing are presented in subsections on 'Intelligent Tools' and on 'Finance'. In addition to the conclusions vis-à-vis the two, hypothesized patterns of concepts in the literature, additional conclusions based on other patterns will be offered.

3.1. Classification language

The classification language after the judicious merger of parts of the *ACM Classification System* with various finance classifications had over 300 distinct concepts, and its hierarchy was six levels deep. At the completion of indexing and classification language refinement, only 141 concepts remained in the classification language. Of the final 141 concepts, 57 were in the Finance sub-tree and 84 in the Intelligent Tools sub-tree. The classification language hierarchy shrank from six to five levels (Rada & You, 1992). The first three levels of the classification language accounted for 25 concepts and are depicted in Fig. 1.

Over 150 concepts were removed from the classification. For instance, the concepts 'theory of simulation' and 'automatic programming' were removed. No article was about either of those concepts. While those concepts are important in the ACM literature, they were not the point of any of the *ESwA* articles.

1 Finance
  1.1 Data Identification
    1.1.1 Methods of Identifying Data
    1.1.2 Types of Data
  1.2 Asset Valuation
    1.2.1 Accounting
    1.2.2 Commodities Valuation
    1.2.3 Corporate Assets
    1.2.4 Derivatives
    1.2.5 Stock Valuation
    1.2.6 Time Value of Money
  1.3 Risk Management
    1.3.1 Dimensions
    1.3.2 Process
2 Intelligent Tools
  2.1 Artificial Intelligence
    2.1.1 Knowledge-Based Methods
    2.1.2 Learning
    2.1.3 Pattern Recognition
    2.1.4 Problem Solving
  2.2 Operations Research
    2.2.1 Discrete Mathematics
    2.2.2 Numerical Analysis
    2.2.3 Probability and Statistics
    2.2.4 Simulation and Modeling

Fig. 1. The top three levels of the classification language are shown in hierarchical format. The numbers represent the hierarchical position. Siblings are sorted alphabetically, except for the causal ordering of the children of 'Finance'.

Examples of adjusting the hierarchy are presented next. The *ACM Classification System* does not have an entry for operations research, but that concept was important enough in the *ESwA* literature to merit a high-level position in the classification. Operations research is an interdisciplinary science that uses methods like mathematical modeling, statistics, and algorithms to support decision-making. The classification language has concepts for mathematical modeling, statistics, and algorithms, and these became the children of 'operation research'. While the notion of decision support is important in *ESwA*, a separate concept in the classification language for decision support was not needed.

The concept of 'Stock Index Valuation' might seem lexically to be in a parent–child, hierarchical relationship with the concept 'Stock Valuation'. However, a 'stock index' is significantly different from a 'stock' for the following two reasons:

- The factors determining the value of a company's stock differ from those needed to determine the value of a stock index. To value a company's stock, one should, among other things, consider the management of the company. In valuing a stock index composed of stocks from hundreds of companies, the management of an individual company is not considered.
- A security whose price is dependent upon or derived from one or more underlying assets is called a derivative,

and the derivative's value is determined by the fluctuations of the underlying asset. The value of a stock index is derived from the value of the stocks in the index.

In the classification language of this paper, the concept of 'Stock Index Valuation' is considered a child of 'Derivatives'.

3.2. Intelligent tools

Of the 65, relevant *ESwA* articles, 29 were from the 1991 to 1996 period, and 36 were from the 2006 to 2007 period. The majority of the articles published between 1991 and 1996 were indexed with an expert systems concept (see Table 1). Only two articles (Coakley, 1995; Jo & Han, 1996) from the 1991 to 1996 period were on a topic in the 'evolutionary computing' sub-tree. For the 2006–2007 articles, the 'Intelligent Tool' indexing was (see Table 2):

- Thirteen times from the 'evolutionary computing' sub-tree and
- Eight times from the 'pattern recognition' sub-tree.

Exactly one article from the 2006 to 2007 period was about 'expert systems' (Kumra, Stein, & Assersohn, 2006). The emphasis in 2006–2007 was on quantitative methods that find patterns in data semi-automatically. Thus, the hypothesis is supported that the field has moved dramatically from considering expert systems approaches to finance to considering evolutionary computation approaches to finance.

Table 1
This concept versus frequency table shows the frequency of use in indexing for those concepts that occurred more than once in the period 1991–1996

Concept	Frequency
2.1.1.2 Expert system	11
2.1.1.2.1 Expert system evaluation	3
2.1.1.2.2 Expert system explanation	3
2.1.1.3 Frames and scripts	2
2.1.1.4 Semantic networks	2
2.1.2.3 Evolutionary computing	2
2.2 Operations research	2

In the 'concept' column appears first the numbering of the concept to indicate the concept's position in the hierarchy and then the name of the concept. Concepts from the 'Expert System' sub-tree were used in indexing 17 times (11 + 3 + 3).

Table 2
This 'concept versus frequency' table shows the frequency of use in indexing for those concepts that occurred more than once in the period 2006–2007

Concept	Frequency
2.1.1.1 Case-based reasoning	2
2.1.2.2 Decision-tree learning	2
2.1.2.3 Evolutionary computing	3
2.1.2.3.1 Genetic algorithms	2
2.1.2.3.3 Neural nets	8
2.1.3 Pattern recognition	2
2.1.3.3.4 Support vector machine	6
2.2.4.1 Model validation and analysis	2

In the 'concept' column appears first the numbering of the concept to indicate the concept's position in the hierarchy and then the name of the concept. Concepts from the 'evolutionary computing' sub-tree were used 3 + 2 + 8 = 13 times. Concepts from the 'pattern recognition' sub-tree were used 2 + 6 = 8 times.

Many other patterns are evident in the data. For instance, case-based reasoning is a method that was established first for natural language processing (Aamodt & Plaza, 1994). However, the method is used in two articles from this study:

- Oh and Kim (2007) apply case-based reasoning to predicting financial market collapse in Korea and
- Chun and Park (2006) apply case-based reasoning to predict the Korea Stock Exchange Index price based on the volume and price of the Index.

While those in natural language processing use case-based reasoning to handle common sense, the *ESwA* authors use case-based reasoning to process large amounts of numeric data.

Much of the classic, knowledge-based work in artificial intelligence does not appear in the recent work on financial applications. For instance, one might expect that natural language processing programs would be used to parse news stories and issue warnings (as when a declaration of war leads to a rise in gold prices). However, not a single article used natural language processing tools. Two articles (Andoh-Baidoo & Osei-Bryson, 2007; Ince & Trafalis, 2006) were about the impact of news stories on corporate stock values but neither employed natural language processing tools.

3.3. Finance

The finance concepts from the 65 articles have the following distribution, in descending order of frequency and stopping at a frequency of four (see Tables 3 and 4):

- Thirty-three articles were indexed with concepts in the 'Accounting' sub-tree,
- Five were indexed with concepts in the 'Data Identification' sub-tree,
- Five were indexed with the concept 'Stock Index',
- Four were indexed with the concept 'Stock Valuation', and
- Four were indexed with concepts in the 'Risk Management' sub-tree.

Table 3
For the literature between 1991 and 1996 this table shows the frequency with which concepts from the 'Finance' tree were used in indexing

Concept	Frequency
1.2.1 Accounting	5
1.2.1.1 Auditing	8
1.2.1.1.1 Bankruptcy	2
1.2.1.1.2 Credit rating	4
1.2.3.1 Product choice	2

Concepts used only once are not listed.

Table 4
For the 2006–2007 literature, the frequency of occurrence of concepts is shown next to the concept from the 'Finance' tree

Concept	Frequency
1.1.1 Methods of identifying data	2
1.1.2.2.2 News	2
1.2 Asset valuation	2
1.2.1.1 Auditing	2
1.2.1.1.1 Bankruptcy	6
1.2.1.1.2 Credit rating	5
1.2.4.3 Stock index	5
1.2.5 Stock valuation	4

Only concepts used more than once are shown.

The preceding list accounts for 51 of the 65 articles. The Accounting sub-tree holds the lion's share of the academic interest in finance. 'Cost Accounting' and 'Auditing' are children of 'Accounting', and 'Bankruptcy' and 'Credit Rating' are children of 'Auditing'.

The hypothesis tested false that the primary concern would be for 'Stock Valuation'. In fact, 'Stock Valuation' as an indexing concept was only applied four times. By contrast, 'Accounting' concepts were used to index 33 of the 65 articles or approximately half the articles.

As one looks deeper for insights about the literature, one notices other patterns. The papers that forecast stock index values rely primarily on the previous prices of the stock index, as in Hyup Roh (2007) and (Oh & Kim, 2007). One might conjecture that the four articles on 'Stock Valuation' would use more than numerical, time series data. However, the results are otherwise. The four papers that address stock value may be described as follows:

- Wu, Lin, and Lin (2006) use the history of stock prices and national money supply in assessing technology stock via decision-tree learning.
- Wang and Chan (2006) use previous stock prices in predicting the future stock prices of individual technology stocks via decision-tree learning.
- Huang, Hsu, and Wang (2007) use price, volume, and 'book to market' values to choose stocks to purchase via pattern recognition.
- Hassan, Nath, and Kirley (2007) use the history of technology stock prices to predict future technology stock prices via evolutionary computing.

While in principle the determination of stock prices should consider subjective factors such as management quality, the research papers that addressed stock valuation employed only data that is readily available in numerical time series. Whether this choice of input data reflects the character of the problem or reflects the choice of tool is unclear. The four methods that were applied to stock price prediction work well with numerical, time series data.

Another set of observations concerns omissions from the literature. The 'Derivatives' concept has three children: 'Futures Prices', 'Options Prices', and 'Stock Market Index'. No articles were about 'Futures Prices' or 'Options Prices', although the futures and options markets are enormous.

Given that a robust, real-world, financial investing system would need to identify data, value assets, and manage risk, one would expect some papers to address these three phases of financial investing. However, no single paper addressed the three phases. Kim (2006) addressed both data identification and asset valuation. A few papers experimented with both asset valuation and risk management, such as Dempster and Leemans (2006) and Oh, Kim, Min, and Lee (2006).

4. Discussion

This 'Discussion' section contains 'Intelligent Tools' and 'Finance' subsections. Each subsection looks at the match between computer tools and financial problems. Gaps in the literature are also noted.

4.1. Intelligent tools

As one can see cycles in the value of financial assets, one can also see cycles in the perceived interestingness of academic topics (as reflected in funding and publications). In the field of artificial intelligence, evolutionary computing was, roughly speaking, popular from 1955 to 1975, expert systems were popular from 1975 to 1995, and evolutionary computation, from 1995 to the date of this paper. This leaves unaddressed the question of whether for a given problem, such as speculative finance, the problem is more appropriately addressed with one tool or another.

The literature on valuing market-traded assets and derivatives uses technical indicators – largely moving averages – and is pattern recognition oriented. A typical pattern recognition approach to a market problem is indicated by Chun and Park (2006). The input data were daily values over five years for five attributes of the Korean Stock Price Index: daily high and low values, daily opening and closing values, and daily trading volume. On the other hand, the financial accounting work looks at fundamental data. For example, the input data for the bond rating work of (Kim & Lee, 1995) considers the quality of management and the quality of financial policies. The classic expert system's approach has a professional interactively answer questions from the system. Through this user interactivity, the system might collect subjective information, such as a company's management quality.

The systematic approach to classification language construction and indexing provides opportunities for numerous other insights. For instance, the *ACM Classification System* has a concept 'Discrete Mathematics' which has children of 'Combinatorics' and 'Graph Theory'. However, the indexing of the *ESwA* literature did not need these concepts. Might any financial problems be suited to discrete mathematics approaches? Portfolio management could be seen as a combinatorial problem. More generally, a researcher might consider whether an unused concept is a marker for a worthwhile research direction.

Vellido et al. (1999) concluded that the literature is rich with neural network applications and suggested that future work explore the combining of knowledge-based techniques with neural network techniques. Such work is now occurring. For instance, Tsakonas, Dounias, Doumpos, and Zopounidis (2006) use logic neural nets that can be understood by people. Genetic programming modifies the architecture of the neural net by adding or deleting nodes of the network in a way that preserves the meaning of the neural net to people and to the net itself. Tsakonas (2004) uses syntactic restrictions on genetic programming that put knowledge into the evolutionary process. Work from another journal also shows the knowledge-based and evolutionary techniques coming together when Bhattacharyya, Pictet, and Zumbach (2002) add semantic constraints to the genetic operators in their application for investing in foreign exchange markets. The classification language produced for this work can be merged with other knowledge and used as part of a knowledge base for guiding change inside an evolutionary computation system for finance.

4.2. Finance

The results on finance were surprising. Investing in stocks was a much less common application than financial accounting. What are attributes of these two different application domains that might account for the extent to which they are studied? Determining whether a corporation is credit worthy has certain static characteristics. The corporation makes an application for a loan and the bank may take its time in deciding what conditions, if any, to offer for a loan. Once the loan is made, its conditions are not subject to ready change. Investing in stocks or financial derivatives is typically a fast moving activity based on a history of prices. Those prices may be volatile and entry and exit from the market may occur any time. The speculative financial investing problem is more of a time series problem than the financial accounting problem is.

The interest in stock investing coincides with the interest in evolutionary computing. Of the 29 articles between 1991 and 1996, only one was about stock. Of the 36 articles between 2006 and 2007, at least, 11 were about asset valuation for an asset that has substantial time series aspects.

The 'market efficiency hypothesis' says that all knowledge about an asset's price is reflected in the price (Thomas, 2003). Thus, when looking for an asset whose price software could forecast, one might be attracted to an asset that had not been extensively studied. Many of the finance concepts were not used in indexing. For instance, the 'Options' and 'Commodities' concepts from the 'Derivatives' tree were not used. In other words, assets such as options and commodities may be ripe for study, if one wants to develop software that could forecast the asset's price.

5. Conclusion

In this section, the method and results of this paper are summarized, and then directions for future research are suggested. Salient patterns in the literature have been identified. An approach to combining knowledge-based and evolutionary methods is suggested.

5.1. Summary

Two hypotheses have been experimentally tested in a literature review. One hypothesis is that the trend is from the use of expert system tools to the use of evolutionary computation tools, and the other hypothesis is that speculative investing is the most popular finance application area. The tests of these hypotheses provide information upon which a model for directions in research on intelligent, financial systems can be built.

All the articles in *ESwA* from 1991 to 1996 and 2006 to 2007 on the topic of finance that were accessible in September 2006 were retrieved. In a combination of top-down and bottom-up approaches, a classification language was built and used to index the articles. The classification language had a 'Finance' tree and an 'Intelligent Tools' tree. Each article was categorized with one concept from the Finance tree and one concept from the Intelligent Tools tree.

The classification language was iteratively constructed, might be used by other researchers, is five levels deep, and includes over one hundred concepts. This author intends to use the classification language inside a knowledge-based system for financial investing.

As hypothesized, the pattern of indexing concepts showed that expert systems dominated the 1991–1996 literature, but that evolutionary computation techniques dominated the 2006–2007 literature. However, the finance hypothesis tested false; namely, both in the early period and in the later period, the application area that most attracted attention was financial accounting and not speculative investing. Many patterns in the literature were revealed that suggested further opportunities for research.

5.2. Directions

The financial markets are human markets and do not follow physical laws that can be captured in a single formula. Rather, market phenomena evolve over time as opportunities to make profits in this zero-sum game depend on the changing strategies of the opponent. Thus, among other things, what is important in the input may change over time. An evolutionary system should be able to evolve its input selection, asset valuation, and portfolio management components.

One might have hypothesized that in the journal literature one would have witnessed a trend to go from focusing on one aspect of the problem (such as what data to collect or how to value an asset) to a holistic approach to the problem. However, this seems only partially to have been the case in the *ESwA* literature, where the work is largely asset valuation both in the 1990s and the 2000s. Though not captured in the indexing, the journal literature shows an increasing respect for an integrated approach that addresses data collection, asset valuation, and portfolio management.

The older ESwA literature directly represented cause-and-effect rules in human-readable knowledge bases. The more recent literature emphasizes pattern recognition. Two approaches are suggested to insert cause-effect knowledge into evolutionary programs for financial investing: (1) consider another asset-class, such as commodity derivatives or (2) transform the financial accounting knowledge into an evolutionary framework.

The first approach exploits an asset (commodity derivatives) not addressed in the *ESwA* literature but for which a wealth of knowledge exists about the relationship between publicly-available data and the asset valuation (Geman, 2005). Commodity prices, of course, depend on supply-and-demand, but unlike corporate stocks, the variables that affect commodity supply-and-demand tend to be published, numerical values. For instance, weather affects the supply of agricultural products and the demand for energy products. Information about weather is readily available and rules relating particular weather patterns to particular supply or demand results are available.

The second approach is to transform the knowledge from the knowledge-based work into a form that can be used by evolutionary systems. For instance, neural logic nets could represent some of the cause-effect knowledge from a bankruptcy system and become part of an evolutionary computation system for predicting stock prices. Some of the bankruptcy variables are readily available online, such as a company's debt, cash flow, and capital assets. The evolution would need to be sensitive to the structure of the rules, as in Tsakonas et al. (2006). Other extensions might:

- make the input selection dynamic with techniques such as those indicated in Kim (2006) and
- add portfolio management by building on the work with betas as in Oh et al. (2006).

Tsakonas (2004) augments genetic programming with syntactic knowledge, but that work can be extended by adding semantic and pragmatic knowledge. The classification language produced for this work incorporates semantic relations that might be of use in supporting genetic operators.

By systematically reviewing the existing literature one sees the gaps in what has been done and the direction forward. Opportunities exist to exploit concepts in finance and artificial intelligence that have not been adequately addressed. A promising research direction is to combine the earlier knowledge-based work on financial accounting with the more recent work on evolutionary computation for financial investing.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Artificial Intelligence Communications, 7*(1), 39–52.

American Heritage Dictionary of the English Language. (2000). 4th ed. Houghton Mifflin Company.

Andoh-Baidoo, F. K., & Osei-Bryson, K.-M. (2007). Exploring the characteristics of Internet security breaches that impact the market value of breached firms. *Expert Systems with Applications, 32*(3), 703–725.

Bauer, R. J. (1994). *Genetic algorithms and investment strategies*. New York: Wiley.

Bhattacharyya, S., Pictet, O. V., & Zumbach, G. (2002). Knowledge-intensive genetic discovery in foreign exchange markets. *IEEE Transactions on Evolutionary Computation, 6*(2), 169–181.

Bodie, Z., & Merton, R. C. (1998). *Finance* (Preliminary ed.). Upper Saddle River, NJ: Prentice Hall.

Chun, S.-H., & Park, Y.-J. (2006). A new hybrid data mining technique using a regression case based reasoning: Application to financial forecasting. *Expert Systems with Applications, 31*(2), 329–336.

Coakley, J. R. (1995). Using pattern analysis methods to supplement attention-directing analytical procedures. *Expert Systems with Applications, 9*(4), 513–528.

Coulter, N., French, J., Glinert, E., Horton, T., Mead, N., Rada, R., et al. (1998). Computing classification system 1998: Current status and future maintenance, report of the CCS update committee. *Computing Reviews, 39*(1), 1–62.

Dempster, M. A. H., & Leemans, V. (2006). An automated FX trading system using adaptive reinforcement learning. *Expert Systems with Applications, 30*(3), 543–552.

Devedzic, V. (1999). A survey of modern knowledge modeling techniques. *Expert Systems with Applications, 17*(4), 275–294.

Dobrev, P., Strupchanska, A., & Toutanova, K. (2001). CGWorld-2001 – New Features and New Directions. In 9th international conference on conceptual structures (pp. 1–11 <http://www.cs.nmsu.edu/~hdp/CGTools/proceedings/papers/CGWorld.pdf>) Stanford University, California.

Elsevier. (2006a). *Elsevier.com – Expert Systems with Applications*, Retrieved November 2006, from <www.elsevier.com/locate/eswa>.

Elsevier. (2006b). *Scopus Info*, Retrieved November, from <www.info.scopus.com>.

EndNote. (2006). Thomson Retrieved November, from <www.endnote.com>.

Ezziane, Z. (2006). Applications of artificial intelligence in bioinformatics: A review. *Expert Systems with Applications, 30*(1), 2–10.

Geman, H. (2005). *Commodities and commodity derivatives: Modeling and pricing for agricultural, metals and energy*. New York: John Wiley & Sons.

Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. *Expert Systems with Applications, 33*(1), 171–180.

Huang, Y.-P., Hsu, C.-C., & Wang, S.-H. (2007). Pattern recognition in time series database: A case study on financial database. *Expert Systems with Applications, 33*(1), 199–205.

Hyup Roh, T. (2007). Forecasting the volatility of stock price index. *Expert Systems with Applications, 33*(4), 916–922.

IEEE. (2006). IEEE Computational Intelligence Society, Retrieved November, from <http://ieee-cis.org/pubs/tec/>.

Ince, H., & Trafalis, T. B. (2006). Kernel methods for short-term portfolio management. *Expert Systems with Applications, 30*(3), 535–542.

Jagerman, E. (2006). *Creating, maintaining and applying quality taxonomies*. Amsterdam, Netherlands: Jagerman via Lulu.

Jo, H., & Han, I. (1996). Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction. *Expert Systems with Applications, 11*(4 SPEC. ISS), 415–422.

Kim, B.-O., & Lee, S. M. (1995). Bond rating expert system for industrial companies. *Expert Systems with Applications, 9*(1), 63–70.

Kim, K.-J. (2006). Artificial neural networks with evolutionary instance selection for financial forecasting. *Expert Systems with Applications, 30*(3), 519–526.

Kingdon, J. (1997). *Intelligent systems and financial forecasting*. London: Springer-Verlag.

Kumra, R., Stein, R. M., & Assersohn, I. (2006). Assessing a knowledge-based approach to commercial loan underwriting. *Expert Systems with Applications, 30*(3), 507–518.

Liao, S. H. (2003). Knowledge management technologies and applications – literature review from 1995 to 2002. *Expert Systems with Applications, 25*(2), 155–164.

Liao, S. H. (2005). Expert system methodologies and applications – A decade review from 1995 to 2004. *Expert Systems with Applications, 28*(1), 93–103.

Mili, H., & Rada, R. (1988). Merging thesauri: Principles and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 10*(2), 204–220.

Nedovic, L., & Devedzic, V. (2002). Expert systems in finance – A cross-section of the field. *Expert Systems with Applications, 23*(1), 49–66.

Oh, K. J., & Kim, T. Y. (2007). Financial market monitoring by case-based reasoning. *Expert Systems with Applications, 32*(3), 789–800.

Oh, K. J., Kim, T. Y., Min, S.-H., & Lee, H. Y. (2006). Portfolio algorithm based on portfolio beta using genetic algorithm. *Expert Systems with Applications, 30*(3), 527–534.

O'Leary, D. E. (1991). Artificial intelligence and expert systems in accounting databases: survey and extensions. *Expert Systems with Applications, 3*(1), 143–152.

Oz, E. (2006). *Management information systems* (5th ed.). Boston, Massachusetts: Thomson.

Rada, R., Mili, H., Letourneau, G., & Johnston, D. (1988). Creating and evaluating entry terms. *Journal of Documentation, 44*(1), 19–41.

Rada, R., & You, G.-N. (1992). Balanced outlines and hypertext. *Journal of Documentation, 48*(1), 20–44.

Thomas, J. D. (2003). *News and Trading Rules*. Unpublished PhD, Carnegie Mellon University, Pittsburgh.

Tsakonas, A. (2004). Towards neural-symbolic integration: The evolutionary neural logic networks. In Intelligent Systems, 2004. Proceedings. 2004 2nd international IEEE conference, Vol. 151 (pp. 156–161).

Tsakonas, A., Dounias, G., Doumpos, M., & Zopounidis, C. (2006). Bankruptcy prediction with neural logic networks by means of grammar-guided genetic programming. *Expert Systems with Applications, 30*(3), 449–461.

Vellido, A., Lisboa, P. J. G., & Vaughan, J. (1999). Neural networks in business: A survey of applications (1992–1998). *Expert Systems with Applications, 17*(1), 51–70.

Waltz, D. (1989). Scientific DataLink's artificial intelligence classification scheme. *Artificial Intelligence Magazine, 6*(1), 58–63.

Wang, J.-L., & Chan, S.-H. (2006). Stock market trading rule discovery using two-layer bias decision tree. *Expert Systems with Applications, 30*(4), 605–611.

Wu, M.-C., Lin, S.-Y., & Lin, C.-H. (2006). An effective application of decision tree to stock trading. *Expert Systems with Applications, 31*(2), 270–274.