Applying the Theory of Semiotics when developing Annual Reports of Organizations: An Approach with Artificial Neural Networks

Marcelo Seido Nagano
drnagano@usp.br
Universidade de São Paulo

Marcelo Botelho Costa Moraes
mbotelhocm@uol.com.br
Universidade de São Paulo

Vinicius Amorim Sobreiro
sobreirovinicius@gmail.com
Universidade de São Paulo

Aplicando a Teoria da Semiótica ao Desenvolver Relatórios Anuais de Organizações: Uma Abordagem com Redes Neurais Artificiais

Abstract

The aim of this paper is to use the Theory of Semiotics to discuss the relationship of language and its various interpretations in accounting. In this study, a Self-Organizing Map Model (SOM) of Artificial Neural Networks is proposed. The purpose of developing this application is to compare the textual code patterns of annual reports with financial information at different times using accounting demonstrations. Moreover, the aim of this method is to establish semiotic relationships in textual message codes showing that there is an historical relationship between accounting information and textual reports. The obtained results are significant, because they indicate that inherent aspects of language hinder its function.

Keywords: Semiotics. ANN. Annual Reports.

Resumo

O objetivo neste artigo é aplicar a Teoria da Semiótica para discutir a relação de idioma e suas várias interpretações na contabilidade. Neste artigo, para tal aplicação, é utilizada uma Rede Neural Artificial do tipo Self-Organizing Map Model – SOM. O propósito de desenvolver esta aplicação é comparar o código textual dos relatórios financeiros anuais ou demonstrações contábeis com informação financeira em diferentes momentos. Dentro desse contexto, a principal vantagem deste método é estabelecer relações sematológicas existentes entre as informações contábeis e os relatórios textuais. Os resultados obtidos são significantes, porque os mesmos indicam que os aspectos inerentes ao idioma dificultam o estabelecimento de relações sematológicas ou da realização de sua função.

Palavras-chave: Semiótica. ANN. Relatórios Anuais.

1 Introduction

Being able to communicate is difficult and this is a universal fact which affects all areas of human knowledge, mainly concerning understanding information to be transmitted. Accounting is a science whereby information is provided to its users to take decisions at various levels.

One of the great dichotomies of accounting refers to the level of specific knowledge that the user of its information should have. On the other hand, one of the main objectives of accounting is to understand the information and therefore it is difficult to draw a line to separate technical terminology from general knowledge. Taking this into consideration, accounting information should reach the largest possible number of users, but this does not necessarily mean losing information in the transformation process or even simplifying the information.

Rev. Tecnol. Fortaleza, v. 31, n. 1, p. 7-14, jun. 2010

7

The main aim of this paper is to study accounting demonstration codes in textual reports using the Theory of Semiotics. Considering this objective, an application of the Self-Organizing Map Model (SOM) is proposed to measure the existence of a semiotic pattern in these codes. As a result, it will be possible to analyze if there is a generalized pattern of codes, considering semiotics in detail used by the editors of annual reports.

This paper is organized as follows: in the first section, the context and relevance of this paper is presented; the second section demonstrates the theoretical foundations of semiotics, communication in accounting, and of ANNs; in the fourth section, the criteria and conditions are shown; the fifth section discusses the obtained results; and finally the last section concludes the paper.

2 The Theory of Semiotics and the Theory of Communication in Accounting

The basic aim of financial reports in accounting, within its strict concept of an information system, is to provide useful information for investment and profit decisions. For the FASB (Financial Accounting Standard Board), accounting information can only be useful if it is relevant, reliable and comparable (KOHONEN, 1989). The theory of accounting is divided to two different classes:

- Theory as Language: accounting is a language which consists of: (I) pragmatics, (II) semantics; and (III) syntax;
- Theory as Reasoning: accounting divides the relationship of assumptions and accounting principles between deductive reasoning and inductive reasoning.

Based on these divisions, the focus of this paper is the Theory as Language, specifically, in the semiotic relationship in terms of the process of accounting communication using Annual Reports. A seguir será feita uma revisão da literatura em publicações que tenham relacionamento simultâneo com energia solar fotovoltaica e a região Nordeste do Brasil. Soltermann e Silva (1998) estudaram a possibilidade de se produzir eletricidade e hidrogênio eletrolítico pelo uso da energia fotovoltaica no Nordeste brasileiro. Eles demonstraram a viabilidade técnica e econômica da proposta quando associada à hidroeletricidade secundária da região. Fraidenraich e Vilela (2000) estudaram o desempenho de sistemas de bombeamento solar fotovoltaico pelo método de utilizabilidade, destacando a adequação do método proposto para os referidos sistemas. De Lima e Veziroglu (2001) modelaram um programa de geração de hidrogênio de origem solar-fotovoltica para o Nordeste brasileiro, eles observaram que a introdução de um programa daquela natureza iria influenciar na matriz energética daquela região, reduzindo poluição ambiental e garantindo melhor índice de qualidade de vida para a população. Tiba (2001), em um trabalho exaustivo, aperfeiçoou as informações existentes sobre radiação solar no Nordeste brasileiro. Bione *et al*. (2004) fizeram a comparação do desempenho de sistemas de bombeamento de água fotovoltaico com seguidores solares e perceberam redução da ordem de 19% nos custos destes sistemas quando comparados com os tradicionais, os quais trabalham de forma fixa.

2.1 The Theory of Communication and Accounting

Accounting is defined as a system of information and evaluation which provides its users with economical, financial, physical and productivity demonstrations and analyses. Moreover, many managers recognize accounting as a language of businesses. However, problems can arise in accounting regarding the transmission of information in demonstrations and analyses.

Accounting, or more precisely accounting language, should be approached in terms of relevance of information to be transmitted. As this relevance of accounting information in finance markets is an area which has been studied in-depth by accounting professionals, this approach is due to the fact that it is relatively easy to carry out interaction tests with information and market agents. The theory of accounting as language can be divided into three language components:

- Pragmatics: This part studies the effects of language. This is usually the main focus in accounting studies;
- Semantics: This part studies the meaning of language observing the specific meaning which transmitted information provides to each one of the receivers; and
- Syntax: This part studies the logic or grammar of the language.

Considering the communication process, some factors make transmitting requested information more efficient. Thus, for individuals involved in the communicative process, the following aspects should be observed: (I) communication abilities in writing and speaking; (II) reading and listening; (III) relationship attitudes with information; and (IV) knowledge levels of the subject.4 Sistemas fotovoltaicos autônomos

2.2 The Problem of Transmitting Accounting Messages

Currently, companies ever increasingly need reappraisal for their communication processes with their public, involving technology and communications channels. Accounting information is subject to various problems of understanding and communication. As well as the problem of previous having to be aware of accounting knowledge, traditional systems of accounting have a high level of semantic meaning. People who do not have this knowledge are less capable of absorbing information (DUNN and GRABSKI, 2000). The need for previous knowledge could be related to the loss of transparency of information caused by accounting symbols that do not reflect reality or meaning (MACINTOSH and SHEARER, 2000), as well as the surplus of summarizing information caused by successive coding stages.

Historically, it can be observed that accounting has been constantly changing in form and content. This change is influenced by developments in culture and in the objectives of the society (MACINTOSH et al., 2000). Widely used Debit/Credit systems are not significant in the semantic sense, not because they are limited to debits/credits, but because this current system based exclusively on this type of information does not reflect all the phenomena of the real world. (DUNN and GRABSKI, 2000). Thus, the problems presented in accounting communication are associated to three stages of the process:

- Noises: in the accounting process, noise problems are mainly associated with technical failures related to the communication channel;
- Feedback: in accounting, there are few ways of obtaining feedback about communication; and
- Coding: this is the main cause of most of the problems of understanding the message transmitted by accounting. Generally, the coding problem is a result of the theory of communication (pragmatic, semantics and syntax).

2.3 Semiotics in Accounting

The importance of the study of Semiotics in Accounting results from problems mainly in coding, because the issuer always informs something to the receptor using a Semiotic pattern which, in this case is Accounting Language. Thus, associating semiotics with the theory of communication, the effect of the relationship symbols-meaning for the user of Accounting information is evident.

In this relationship or model, companies interpret economical transactions (Semantic) and register them according to accounting norms (Syntax), encoding it in terms of accounting demonstrations (symbols) which are available to the users (Observers). These accounting demonstrations bring about interactions between the users and the organizations. The interactions help the users of Accounting Information to take decisions (Pragmatic).

In spite of the limitations of the presented model, the relationship of semiotics in accounting is not that simple. It is often the case that the signs do not arouse the same perception in different observers, causing communication breakdowns. This can be observed in more detail in some accounting concepts such as, "The True and Fair View", which can be considered as "fair and true" in accounting in France and is relatively different in Japan.

The semiotic issue of accounting does not observe only international semantic relationships, even different groups of users can present different interpretations of the same symbols. Semiotics in accounting is of extreme importance for the efficiency of the communication process. However, this theory is extremely complex and there are no methods which can evaluate the efficacy of the communication process or simply of the semiotics involved.

Taking this into account, this paper proposes an application of ANN modeling which attempts to gauge the semiotic relationship between textual reports and accounting demonstrations.

3 Artificial Neural Networks

ANNs are support techniques for decisions. They are based on mathematical simulations that resemble the logic of human reasoning. ANNs are obtained by modeling relevant variables in a certain analysis. Historically, ANNs have appeared in mathematical and statistical studies, involving areas such as biology, mathematics, psychology and computing science.

Due to the fact that ANNs are dynamic, these networks can self adjust themselves for certain functions. These functions are used essentially to predict changeable environments. ANNs have been used in situations considered problematic such as: (I) problems that do not have exact models; and (II) in environments that have excessive changes. The model was based on neuroanatomy, artificially simulating a communication system in brain cells and based on fuzzy logic. In this model, the cells or the neurons are interlinked in networks whereby each unit receives a series of inputs and joins them in a single output.

3.1 Artificial Neurons

Based on neuroanatomy, the model artificially stimulates connections between the neurons in a network, whereby each computational unit (neuron) receives a series of inputs and joins them, making a new input and a series of subsequent neurons for the final output of the network, which should be the final solution to the problem. Developed in 1943 by McCulloch and Pitts, the Perceptron is a mathematical model of how a biological neuron works. Thus, the neuron can be defined by the following model shown in Figure 1.
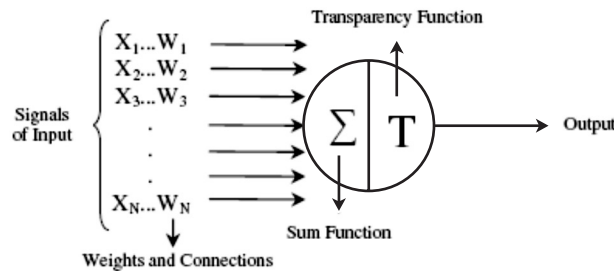

Figure 1: Model of Artificial Neuron.

Each neuron (j) is numbered and indexed (in an analogy of the biological neuron there are various inputs (Dendrites)) and is powered by the respective weights wj (Synaptic Connections) and with the sum function simulating the capture of stimulus received, it accumulates the signal received by the neuron in sj.

$$s_j = \sum_{i=1}^{n} x_i w_{ji} \tag{1}$$

From the value of the weighted sum, the transparency function *(T)* uses the threshold θ to determine the output yj:

$$y_j = f(s_j) = \begin{cases} 1 & s_j \geq \theta \\ 0 & s_j < \theta \end{cases} \tag{2}$$

The output yj is determined by the function f(sj), usually a sigmoid function, for example, the logistics function (HAYKIN, 1994). Thus, the output generated by the neuron will be one case if the considered sum is higher or equal to θ and zero if less.

3.2. Self-Organizing Model

Kohonen proposes the Self-Organizing Map Model (SOM) that classifies various elements which present similar characteristics in clusters. This model of ANNs is a non-supervised learning model. (FAUSETT, 1994). The SOM is formed by two layer units: one input layer and one competitive layer, with n access and m possible clusters. (KOHONEN, 1989).

In this way of competitive learning, each neuron, regarding a certain cluster, has a vector of weights which can be compared to the access standard, and the neuron presents a smaller difference between the vector and the chosen standard as a winner. The following formula is used to update the weight vectors in the n neurons:

$$w_j(n+1) = \begin{cases} j \in \Lambda_{i(x)}(n), & w_j(n) + \eta(n)[x(n) - w_j(n)] \\ j \notin \Lambda_{i(x)}(n), & w_j(n) \end{cases} \tag{3}$$

Where:

$w_j(n)$ is the weight of the neuron;

$\eta(n)$ is the learning rate; and

$\Lambda_{i(x)}$ is the function neighborhood centered around the "winner" neuron $i(x)$.

The $\eta(n)$ and $\Lambda_{i(x)}(n)$ are altered dynamically during the learning process so that better results can be attained. Therefore, the success of forming the clusters depends on how the parameters of the algorithm are adjusted. Unfortunately, no theoretical basis to select these values could be found in the literature. These values are determined by a trial and error process. (HAYKIN, 1994). Regarding clustering characteristics and non-supervised learning, Kohonen´s model is very efficient in terms of creating clusters.

4. Method

The development of a model to gauge semiotics in accounting is something of extreme complexity, because there are few modeling techniques which can work with data.

In order to develop the model, the following was selected: one aspect of semiotics and the coding of numerical information from accounting demonstrations in textual information analyzed in textual reports. This information comes from Annual Reports which is widely published by open capital companies, addressed to investors and the financial market.

The objective of the method is to observe or identify if there is a semiotic standard in the process of textual coding, once the premise is adopted that the numerical information of accounting demonstrations does not represent the reality to be transmitted. In spite of this, the method is not invalid, because the accounting demonstrations present the textual information numerically.

Furthermore, the Theory of Measuring Information proposes information techniques which are based on statistical models and cybernetics. (CROPLEY, 1998). Taking this into account, the SOM model presents all the necessary characteristics, and is self-organized, not linear and can generalize statistical characteristics in its models. Therefore, this model is appropriate for this application.

The SOM model is self-organized and does not need examples to learn. It can identify common patterns in data and also divide these standards into clusters that present similar characteristics. A cluster is defined as an enclosed structure of properties, isolated externally and united internally. (MANGIAMELI et al., 1996).

The model was applied as follows. Firstly, ANNs are used to detect the pattern of numerical information from the accounting demonstrations which form groups made by companies with similar financial characteristics. Secondly, the same model is applied to obtain textual information from the reports that form the same number of groups. The Semiotics of the relationship number vs. text is in the coding pattern, in other words, the groups formed by ANNs for numerical information and for textual information are similar.

A similar application was developed by (BACK et al., 2001), however these authors did not come to a conclusion which demonstrates the relationship between textual and numerical information. This could have happened due to the

approach which did not observe two aspects: (I) the semiotic theory was not used, because the objective was the simple comparison of standards; and (2) the main factor, the sample was formed by companies from various countries (the United States, Sweden, Finland, Norway, Canada, Austria, Germany, Italy, Holland, Portugal, the United Kingdom, France, Spain and Switzerland) which use different accounting standards (Generally Accepted Accounting Principles - GAAP), making the comparison impossible.

To eliminate the problem of accounting and language pattern when developing the method, the database of the United States Securities and Exchange Commission (SEC) was used. This database presents all the complete Annual Reports of all the open capital companies in the United States stock exchange. The reports are made available in the10-K form, where general information can be found concerning the business of the company and its financial standing. From this information found in form 10-K, some information related to the financial situation of companies was selected.

To obtain numerical accounting information, the Economática database was used as the accounting demonstrations were already standardized. In the sample, 34 companies were selected for three consecutive years: 2000, 2001 and 2002.

These companies are registered in the 2834 code of Pharmaceutical Preparations according to the classification of the Standard Industrial Classification Code (SIC) which divides companies into area of performance. Although the sample focuses on pharmaceutical companies, the method can be applied to any other sector or country.

In order to develop the SOM model of financial classification, information from the balance sheet, end of year result demonstration, and the cash flow demonstration were used. Each paragraph of the text of the accounting demonstrations was considered as a variable up to the limit of twenty five variables, because four companies of the sample present more than twenty five variables. To use textual variables in ANNs, the text was transformed by simplifying the ASCII system which allocated a numerical value to each alphabetical symbol. To simplify, text signal such as full stops and commas were eliminated from the text, as well as the spaces between words.

5. Analysis of the Result

The results obtained by clustering 34 companies are analyzed considering three aspects:

- The annual (2002, 2001, 2000) comparison of the level of adaptation of textual information regarding financial information, measured by the success percentage when comparing groups;
- The evolutionary (2000-2002) comparison of the cluster of the same type of information, indicating the cohesion in the information regarding 2002 in relation to the same data of 2001 and 2000; and
- In each type of information, the recurrent grouping of companies, in other words, the data of the same company in different years.

5.1. Results of the Financial SOM and Textual SOM in 2000, 2001 and 2002.

Table 1: Financial SOM and Textual SOM.

CLUSTER	CLUSTERS 2000		CLUSTERS 2001		CLUSTERS 2002	
	FINANCIAL	TEXTUAL	FINANCIAL	TEXTUAL	FINANCIAL	TEXTUAL
Cluster 1	20	11	3	4	3	11
Cluster 2	-	5	-	-	-	-
Cluster 3	6	-	3	6	1	5
Cluster 4	-	5	-	-	-	-
Cluster 5	2	1	2	6	3	2
Cluster 6	-	-	-	-	-	-
Cluster 7	3	-	4	3	4	1
Cluster 8	-	-	-	-	-	-
Cluster 9	3	12	21	14	20	12
SUCCESS	23.53%		42.42%		41.94%	

In 2002, only 23.53% of the companies were clustered in a textual way. This first analysis indicates that standardization does not exist in the companies when coding numerical information in the text. In the other cases, regarding 2000 and 2001, the accordance percentage among the financial and textual grouping is significantly larger, close to 42%, which indicates a larger standardization of coding, but in general terms this percentage is still very low. Table 2 shows the development of the Financial SOM and Textual SOM.

Table 2: Evolution of Financial SOM and Textual SOM

EVOLUTION		
SOM	**YEAR**	**2002**
Financial	2001	81.82
	2000	75.00
Textual	2001	45.45
	2000	37.50

In this case, classification is compared, made by the SOM Financial network, for the three years using 2002 as a basis, because this year is the most recent. In the case of the evolution of textual grouping, the success percentage in 2002 is not so high. This could be due to the fact that the performance in 2002 (of only 23.53%) in the textual classification is less than the results from 2001 and 2000. Another relevant point is in the variation of the comparison 2002-2001 and 2002-2000, because the difference of approximately 7% both for the financial and textual network demonstrates a tendency of similar development among the financial and textual standards, although this presents a low relationship. In the case of the general grouping, observed in the grouping of the three years of the companies, the information is presented in Table 3.

Table 3: General SOM

	CLUSTERS	
SOM	FINANCIAL	TEXTUAL
SUCCESS	66,67%	51,52%

In this network, a certain clustering level in the same companies in different years can be observed. This is due to the high cohesion of information. In other words, the financial information from 2002 is highly dependent on the financial information from the previous years. In the case of textual information, this phenomenon also happens, indicating that this information follows a similar tendency to the numerical information.

6. Final Considerations

Accounting as a language has communication breakdowns. Thus, communication studies and semiotics in this area of knowledge is of vital importance. These studies should solve or understand communication problems.

The obtained results demonstrate that the standards of financial information, here considered as absolute, do not follow textual information to the rule. One of the main problems is the fact that unfortunately there are many names for one concept in all areas as well as different concepts for only one word. However, this paper presents some positive characteristics. Although there is a low relationship among the financial and textual information observed (23.53% in 2002 and approximately 42% in 2001 and 2000) the analysis of the evolution of companies in the various clusters indicates a tendency and a very similar proportional variation. In other words, the vectors of numerical and textual information do not follow the same standard, but historically (2000-2002) they tend to have continuity both in financial and textual information, showing a high cohesion.

Therefore, the results indicate that in spite of the fact that companies do not use the same semiotic standard, the same company tends to maintain its semiotic standard when compiling its Annual Reports along over time.

Marcelo Seido Nagano, Marcelo Botelho Costa Moraes e Vinicius Amorim Sobreiro

Future studies will better be able to demonstrate this relationship, mainly if better processing and transformation techniques of the textual information are adopted.

Another way of developing knowledge in the semiotic relationship in coding can be made by altering the focus of an annual sample of various companies for a temporary period of one company, because as the results have demonstrated, there is both strong historical and textual cohesion. Therefore, ANNs can do the coding. In other words, the development of an analyzer of accounting demonstrations is possible as it is trained on the basis of a single semiotic standard.

Therefore, ANNs can be characterized as a strong tool when detecting standards that help semiotic relationship studies in accounting, as the used neural model's analogy is its own system of understanding semiotic human.

References

BACK, B. et al. Comparing numerical data and text information from annual reports using self-organizing maps. *International Journal of Accounting Information Systems*, Amsterdam, n. 2, p. 249-269, 2001.

CROPLEY, D. H. Towards formulating a semiotic theory of measurement information: part 1. *Measurement*, London, v. 24, p. 237-248, 1998.

DUNN, C. L.; GRABSKI, S. V. Perceived semantic expressiveness of accounting systems and task accuracy effects. *International Journal of Accounting Information Systems*, Amsterdam, n. 1, p. 79-87, 2000.

FAUSETT, L. *Fundamentals of neural networks*: architectures, algorithms, and applications. Englewood-Cliffs: Prentice-Hall, 1994.

HAYKIN, S. *Neural networks*: a comprehensive foundation. New Jersey: Prentice Hall, 1994.

KOHONEN, T. *Self-organization and associative memory*. New York : Springer-Verlag, 1989.

MACINTOSH, N. B.; SHEARER, T. The accounting profession today: a poststructuralist critique. *Critical Perspectives on Accounting,* London, v. 11, p. 607-626, 2000.

MACINTOSH, N. B. et al. Accounting as simulacrum and hyperreality: perspectives on income and capital. *Accounting, Organizations and Society,* London, v. 25, p. 13-50, 2000.

MANGIAMELI, P.; CHEN, S. K.; WEST, D. A comparison of SOM neural network and hierarchical clustering methods. *European Journal of Operational Research*, London, v. 93, p. 402-417, 1996.

AUTHORS

Marcelo Seido Nagano

He has a Ph.D. in Mechanics Engineering from the Engineering School of São Carlos (EESC) of University of São Paulo. He is currently Assistant Professor of the Production Engineering Department at the Engineering School of São Carlos (EESC), University of São Paulo (USP). His areas of interest include operational research, accounting, and finance.

Marcelo Botelho Costa Moraes

He has a Master's degree in Production Engineering from the Engineering School of São Carlos (EESC), the University of São Paulo (USP). He has a Bachelor's degree in Accounting from University of São Paulo. Areas of interest include accounting and finance in companies.

Vinicius Amorim Sobreiro

He has a Master's degree in Production Engineering from the Engineering School of São Carlos (EESC), the University of São Paulo (USP). He has a Bachelor's degree in Economics from Toledo University. Areas of interest include management of cost, accounting and finance in companies.e pós-doutorado pela Univesity of New South Wales (1994). Professor Titular pela Universidade Estadual do Ceará. Tem experiência na área de Engenharia Elétrica, com ênfase em Sistemas Elétricos de Potência, atuando em Efeito de Estufa, Saldo de Energia, Energia Solar, Sistemas Solares Térmicos, Dióxido de Carbono.