

## Mensuração da Semiótica na Codificação das Demonstrações Contábeis por Meio de Redes Neurais

MARCELO BOTELHO DA COSTA MORAES  
UNIVERSIDADE DE SÃO PAULO

MARCELO SEIDO NAGANO  
UNIVERSIDADE DE SÃO PAULO

EDGARD MONFORTE MERLO  
UNIVERSIDADE DE SÃO PAULO

Resumo: A contabilidade como uma linguagem de comunicação possui diversos problemas de compreensibilidade na transmissão de sua informação, principalmente com relação à interpretação de conceitos. Para analisar melhor este problema o presente trabalho utiliza a teoria semiótica para discutir a relação da linguagem e suas diversas interpretações e propõe uma metodologia de aplicação do modelo auto-organizado de redes neurais artificiais (*Self-Organizing Feature Maps*) no intuito de comparar os padrões de codificação textual dos relatórios anuais com a informação financeira numérica existente nos mesmos, a fim de mensurar a relação semiótica na codificação da mensagem textual. Assim, utilizando-se como base de dados informações numéricas (demonstrações contábeis) e textuais (relatórios anuais) de empresas do setor farmacêutico dos Estados Unidos aplica-se o modelo auto-organizado para agrupar as empresas que possuem informações numéricas e textuais de padrões semelhantes. Os resultados demonstrados indicam interessante relacionamento entre a informação textual e a informação numérica, principalmente com relação a coesão das terminologias utilizadas pelas empresas, com isso validando a metodologia aplicada e sugerindo possibilidades para futuros estudo que possam elucidar sobre a relação da comunicação e semiótica na contabilidade.

### Introdução

O problema da comunicação é universal e atinge todas as áreas de conhecimento humano, principalmente no tocante a compreensibilidade da informação a ser transmitida. Assim, sendo a Contabilidade uma ciência que tem por objetivo fundamental prover informação à tomada de decisão em diversos níveis, tanto interna quanto externamente às instituições, pode ser vista como ferramenta de gestão e principal linguagem de comunicação dos fatos econômicos e da influência direta nas atividades empresariais.

Para um maior proveito da informação contábil algumas características básicas são necessárias, tais como a *comparabilidade, objetividade, compreensibilidade, oportunidade e relação favorável entre o custo e os benefícios dela esperados* (DIAS FILHO, 2001).

O *Statement of Financial Accounting Concepts n.º 2 (SFAC n.º 2)*, editado em 1980 pelo *Financial Accounting Standards Board (FASB)* determina que “... os benefícios da informação podem ser aumentados se a tornarmos mais compreensível e, assim, útil para um círculo maior de usuários ...” (DIAS FILHO, 2001), mostrando assim que a informação contábil deve atingir o maior número de usuários possível, mas isso não deve significar necessariamente perda de informação durante o processo de transformação ou mesmo simplificação da informação.

Mesmo assim, com parte dos Relatórios Anuais sendo transmitida em forma textual a compreensão da informação contábil não está garantida, pelo contrário, como esses relatórios focam suas informações aos usuários que possuem conhecimentos contábeis reduzidos e os relatórios são confeccionados por Contadores, ou por pessoal da área contábil, podem ocorrer falhas no processo de comunicação.

Nesse aspecto, a Teoria da Comunicação estabelece como processo comunicativo o fluxo de informação partindo do *emissor* – nesse caso o redator do relatório que codifica a informação contábil numérica em textual – enviando a *mensagem* (informação contábil) ao *receptor* – o usuário da informação – através de um *canal* (relatório anual), por meio de um *código*, sujeito a interferências (*ruídos*) no envio da mensagem. O problema seria simples de resolver se não fosse o fato do *código* apresentar certas inconsistências. Isso ocorre, pois, apesar de não existir problema de incoerência semântica na Língua Portuguesa, ou em qualquer outra língua moderna, ou mesmo ambigüidade na terminologia contábil, ao se codificar uma informação com um foco tão abrangente e sendo o codificador um conhecedor de ambos os códigos pode haver certa confusão de sua parte.

O presente trabalho tem como objetivo o estudo da codificação das demonstrações contábeis em relatórios textuais utilizando a Teoria Semiótica e propondo um modelo baseado em redes neurais artificiais para mensurar se existe um padrão semiótico nessa codificação. Com isso será possível analisar detalhadamente se dentro da semiótica utilizada pelos redatores dos relatórios anuais, existe um padrão generalizado de codificação.

## **1. Semiótica e Teoria da Comunicação na Contabilidade**

A contabilidade, dentro de seu conceito estrito de sistema de informação possui como objetivo básico dos relatórios financeiros prover informações úteis para decisões de investimento e empréstimos, no enfoque do usuário externo utilizado neste trabalho. Para o FASB a informação contábil só pode ser útil se for relevante, confiável e comparável (HORNGREN, HARRISON, ROBINSON, 1996).

### **1.1. Teoria da Comunicação e a Contabilidade**

A Contabilidade, definida como um sistema de informação e avaliação que tem por objetivo prover a seus usuários demonstrações e análises de caráter econômico, financeiro, físico e de produtividade com relação à entidade objeto (IUDÍCIBUS et al., 2000), ciência que entre suas diversas atribuições, possui como característica principal o seu reconhecimento por muitos como “linguagem dos negócios” (DIAS FILHO, 2001) apresenta diversos problemas com relação à transmissão da informação contida em suas demonstrações e análises.

Ainda sobre essa ótica, a linguagem contábil – como qualquer outra forma de linguagem – deve ser abordada sob um aspecto de relevância da informação a ser transmitida. Sendo que esta relevância da informação contábil nos mercados financeiros é uma das áreas mais estudadas pela academia de contabilidade, isto ocorre devido à facilidade da realização de testes da interação entre a informação e os agentes do mercado (LOPES, 2002).

A Teoria da Contabilidade como uma Linguagem deve ter destacado o papel de cada um dos três componentes da linguagem (HENDRIKSEN and VAN BREDA, 1999):

- A *Pragmática* – estudo do efeito da linguagem, costuma ser o foco principal dos estudos em contabilidade, uma vez que se a informação contábil desempenha seu papel de informar, o resultado da função pragmática será observado na resposta do receptor;
- A *Semântica* – estudo do significado da linguagem, tem por objetivo observar o significado específico que a informação transmitida gera a cada um dos receptores, estabelecendo a relação semiótica da comunicação;
- A *Sintaxe* – estudo da lógica ou gramática da linguagem, detém-se quase que exclusivamente de aspectos mais técnicos da comunicação, para que esta possua a estrutura necessária à compreensão.

Além dos componentes da linguagem o processo de comunicação é formado por uma série de elementos, que juntos demonstram todo o fluxo da comunicação (figura 1).

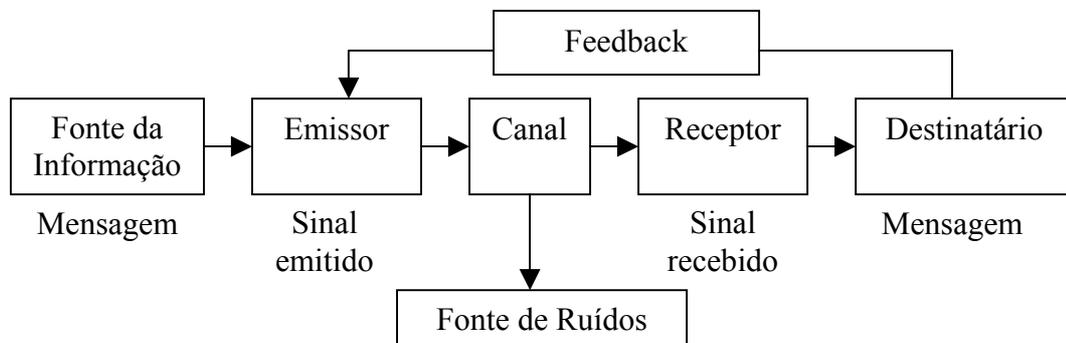


Figura 1 – Esquema geral da Comunicação de Shannon e Weaver (*adaptado* DIAS FILHO, 2001).

De importância para presente estudo, no esquema conceitual acima, a mensagem pode estar sujeita a ruídos, quando elementos são adicionados ao sinal de mensagem gerando perda ou diminuindo a eficiência do processo de comunicação.

Outro fator de extrema importância ao processo é o código, pois este desempenha papel fundamental na elaboração da mensagem. Tendo por objetivo garantir que as representações produzidas sejam compreensivas e universais aos envolvidos no processo da comunicação (NEIVA JR. *Apud* DIAS FILHO, 2001).

## 1.2. O Processo de Comunicação Contábil

O processo de comunicação contábil atende todos os requisitos demonstrados no processo geral de comunicação. Mas com algumas adaptações, mesmo porque a contabilidade é formada por diversos processos de comunicação que ocorrem em seqüência.

Em um modelo geral da comunicação contábil voltada ao usuário externo é possível abstrair o seguinte fluxo: partindo da ocorrência do Evento Econômico (fonte da informação) o emissor (contador ou funcionário treinado para isso) codifica a mensagem – nesse primeiro processo a codificação se dá no lançamento da partida dobrada (débito/crédito) referente ao

evento, disponibilizando a mensagem pelo canal (livros contábeis ou sistemas informatizados) ao receptor (o departamento de contabilidade). No segundo estágio do processo – Demonstrações Contábeis – a fonte de informação se torna o departamento de contabilidade, por meio de seus livros e/ou sistemas, o emissor (nesse caso essencialmente contadores) codifica a informação novamente para sintetizá-la e a disponibiliza pelo canal (Demonstrações Contábeis – Balanço Patrimonial, Demonstração de Resultado do Exercício, DOAR, DFC e outras) para o receptor (usuários dessa informação, normalmente acionistas, investidores e credores). Atualmente o processo ganhou um terceiro estágio, principalmente em companhias de capital aberto, sendo conhecido como os Relatórios Anuais (*Annual Reports*) onde as demonstrações geradas na codificação anterior se tornam uma nova fonte de informação e o emissor - podendo tanto ser um contador como um profissional da área de finanças – o qual promove uma análise textual da situação financeira da empresa.

### 1.3. O Problema da Transmissão da Mensagem Contábil

A informação contábil é passível de diversos problemas de compreensão e comunicação. Além do problema da necessidade de conhecimento contábil prévio, pois os sistemas tradicionais de contabilidade possuem um alto grau de significado semântico, através de tabelas (demonstrações), codificações contábeis e organização da informação. Pessoas que não possuem esse conhecimento têm sua capacidade de absorção da informação extremamente reduzida (DUNN e GRABSKI, 2000).

A necessidade de conhecimento prévio pode estar relacionado com a perda da transparência das informações, causadas por signos contábeis que não reflitam a realidade (MACINTOSH *and* SHEARER, 2000). Além do excesso de sintetização das informações ocasionada por sucessivas etapas de codificação.

Também no tocante a forma, a contabilidade apresenta problemas de compreensão. Sistemas de registro de Débito/Crédito amplamente utilizados são apontados como não sendo semanticamente expressivos, não que um sistema semanticamente expressivo não possa conter débitos e créditos, mas o sistema atual que se baseia exclusivamente neste tipo de informação não reflete fenômenos do mundo real (DUNN *and* GRABSKI, 2000).

Assim, problemas apresentados na comunicação contábil estão associados a três etapas do processo:

**Ruídos:** no processo contábil problemas de ruído são associados essencialmente a falhas técnicas referentes ao canal, como problemas nos sistemas informatizados ou na estrutura tomada como canal, referente a pouca abrangência da informação.

**FeedBack:** na contabilidade existem poucas formas de medir o *feedback* sobre a comunicação em si, mesmo com estudos que demonstram a relevância da informação contábil, principalmente no mercado de capitais onde é possível observar a relação informação *vs.* preço. Ainda assim, é muito difícil verificar quais etapas do processo estão sendo falhas.

**Codificação:** gerador da maior parte dos problemas de compreensão da mensagem transmitida pela contabilidade o problema da codificação recai sobre a teoria da comunicação (pragmática, semântica e sintaxe). Em grande parte o sucesso da codificação está na capacidade do emissor em interpretar os eventos e codificá-los de uma maneira que o receptor ao inverter o processo e decodificar a mensagem obtenha a mesma interpretação. Para tanto é necessária à correspondência entre realidade e mensagem, e os possíveis significados da

mensagem entre o emissor e receptor sejam os mesmos (DIAS FILHO, 2001), para isso é abordada a semiótica na Contabilidade.

#### **1.4. Semiótica na Contabilidade**

A semiótica revela a importância de um agente como sendo o responsável pela existência de um signo e seu significado (LIEBENAU *and* HARINDRANATH, 2002). Assim, o objetivo da semiótica é relacionar qual significado um determinado signo – compreenda-se por signo uma representação gráfica (imagem, desenho), ou textual (palavras) – é capaz de gerar no agente que o observa (CROPLEY, 1998a). Daí a importância do estudo da semiótica na contabilidade, principalmente na codificação, pois se o emissor, seja ele o contador ou a própria empresa num âmbito maior, deseja informar ao receptor, qualquer que seja o usuário (investidor, credor, administração ou governo), é necessário que se utilize um padrão semiótico comum para a mensagem.

Nessa relação empresas interpretam os eventos econômicos (semântica) e os registram de acordo com as normas de contabilidade (sintaxe), codificando os registros contábeis sob a forma das Demonstrações Contábeis (signos) que são disponibilizadas aos usuários (observadores) e que motivam estes a realizarem interações com as empresas, auxiliando a sua tomada de decisão (pragmática).

Entretanto, muitas vezes os signos não despertam a mesma percepção em diferentes observadores, causando falhas na comunicação. Isso pode ser observado mais detalhadamente em alguns conceitos contábeis, como os conceitos de “Justo e Verdadeiro” (conceito internacional de “*The True and Fair View*”), o que pode ser considerado como “justo e verdadeiro” na contabilidade na França é relativamente diferente no Japão e ambos são diferentes do conceito alemão, mas as diferenças internacionais explicam apenas uma parte das diferenças semióticas entre países, existem questões históricas e culturais que regem a diferenciação deste conceito (CUNNINGHAM, 2003).

A questão semiótica na contabilidade não se atém a relações semânticas internacionais, mesmo diferentes grupos de usuários podem apresentar diferentes interpretações de um mesmo signo. Isso ocorreu com o modelo ROI (*return on investment*) que passou a ser utilizado para desenvolvimento de políticas internas da empresa. Com isso transformando este estimador de investimentos em um instrumento administrativo de motivação para os administradores de operações, mostrando que cada divisão poderia influir no resultado global da companhia (LABOURET, 2001).

Portanto, a semiótica na contabilidade é de extrema importância para a eficiência do processo de comunicação. Será apresentada a seguir uma modelagem neural (inteligência artificial) que busca medir a relação semiótica entre relatórios e demonstrações contábeis.

## **2. Redes Neurais Artificiais**

As Redes Neurais Artificiais (RNA's) podem ser caracterizadas como uma das mais avançadas tecnologias de suporte e apoio à decisão. Baseada em simulações matemáticas que se assemelham à lógica de raciocínio humano, as RNA's são obtidas pela modelagem das variáveis relevantes em uma determinada análise e apresentam suas origens há cerca de 50 anos atrás em estudos matemático-estatísticos, envolvendo áreas como biologia, matemática,

psicologia e ciência da computação, mas só obtiveram grande interesse a partir dos anos 80 com o avanço e barateamento dos sistemas de computação e do conhecimento sobre os próprios neurônios (RUMELHART, 1994).

## 2.1. Desenvolvimento da RNA

Para o desenvolvimento de uma RNA são levados em consideração muitos fatores, é necessário que seja feita uma análise completa do setor ao qual se deseja obter os resultados. Assim, uma RNA possui quatro etapas de implementação:

- **Definição da Rede:** processo pelo qual a RNA é desenvolvida teoricamente. Note que se esta modelagem de dados não for realizada de maneira bem estruturada e com analistas de conhecimento, ela pode degenerar toda a rede;
- **Treinamento:** esta etapa é de suma importância à validação da rede, pois é nela que a rede recebe os dados para treinamento;
- **Utilização da RNA:** Após treinamento a RNA passa a fornecer dados confiáveis, utilizando-se então de estimativas. Pode-se entrar com novos dados para se verificar como certa alteração teria efeito sobre determinada função, ou qualquer outro resultado para o qual a rede tenha sido treinada;
- **Manutenção:** A partir de grandes mudanças de ambiente, ou mesmo de variáveis de alta relevância, a rede deve receber certa manutenção para que os pesos das conexões não fiquem desatualizados.

Note que estas quatro etapas, apesar de independentes, estão em um contínuo ciclo e se alteram mutuamente para correções, mantendo a rede sempre bem ajustada (TAFNER, 1995).

Além disso, o desenvolvimento das RNA's pode ser feito por diferentes métodos práticos, através de linguagem de programação e ferramentas de desenvolvimento específicas, além do método misto utilizando-se de ferramentas e linguagens (MORAES e NAGANO, 2001).

## 2.2. Neurônio Artificial

O modelo, baseado na neuroanatomia, simula artificialmente as conexões entre os neurônios, dispostos em uma rede, onde cada unidade computacional (neurônio) recebe uma série de entradas e as combina, gerando uma nova entrada a uma série de neurônios posteriores até a saída final da rede, que deve ser a resposta final ao problema. Desenvolvido em 1943 por McCulloch e Pitts o perceptron é um modelo matemático do funcionamento de um neurônio biológico. Assim, o neurônio pode ser definido pelo seguinte modelo (figura 2):

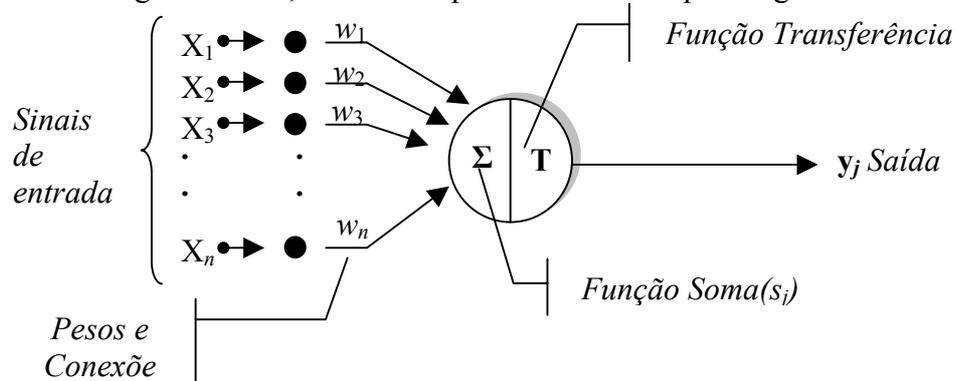


Figura 2 – Modelo de um neurônio artificial.

Cada neurônio ( $j$ ) é numerado e indexado, numa analogia ao neurônio biológico apresenta diversas entradas (dendritos), potencializadas pelos respectivos pesos  $w_j$  (conexões sinápticas) e com a função soma simulando a captação do estímulo recebidos, acumulando de maneira ponderada os sinais recebidos pelo neurônio em  $s_j$ :

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

A partir do valor da soma ponderada, a função de transferência (T) utiliza-se de um limiar  $\theta$  para determinar a saída  $y_j$ :

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

Onde a saída  $y_j$  é determinada pela função  $f(s_j)$ , normalmente uma função sigmóide, como por exemplo, a função logística (HAYKIN, 1994).

Assim, a saída gerada pelo neurônio será 1 se a soma ponderada for superior ou igual a  $\theta$  e 0 se inferior. Onde pode-se comparar  $\theta$  ao limiar de ativação presente nos neurônios biológicos.

O potencial de ação de um neurônio biológico acontece ou não, dependendo da grandeza do estímulo que é aplicado a ele. O limiar é definido como o valor mínimo do potencial de membrana em que vai ocorrer o potencial de ação. Todos os valores de potencial de membrana inferiores ao valor necessário para a produção de um potencial de ação são ditos subliminares. Caso um estímulo não consiga deslocar o potencial de membrana até o valor limiar, a resposta parcial, subliminar, desaparece. O neurônio artificial trabalha de maneira semelhante, onde o limiar de disparo é operado pela função de transferência.

### 2.3. Modelos Auto-Organizados

O modelo de mapas de auto-organização (*Self-Organizing Feature Maps* -SOFM's) de Kohonen propõe uma estrutura topológica que designa cada caso entre as unidades de um conjunto (*cluster*), ou seja, classifica, nesse caso, as empresas entre diversos grupos que apresentam características similares, esta função de agrupamento é observada no cérebro humano, mas em nenhum outro modelo topológico de RNA (FAUSETT, 1994), sendo um modelo de aprendizado não supervisionado.

O SOFM consiste em duas unidades de camada: Uma camada unidimensional de entrada e uma camada bidimensional competitiva, com  $n$  sinais de entrada e possíveis  $m$  *clusters* (KOHONEN, 1989).

Nessa forma de aprendizado competitivo, cada neurônio, referente a determinado *cluster*, possui um vetor de pesos que é comparado com o padrão de entrada, sendo o neurônio que apresentar menor diferença entre o vetor e o padrão escolhido como vencedor (*winner*). Sendo que o “vencedor” e seus “vizinhos” têm seus pesos atualizados (FAUSETT, 1994).

Sendo a seguinte fórmula utilizada para atualização dos vetores de peso nos  $n$  neurônios:

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

Onde,

- $w_j(n)$  é o peso do neurônio  $j$ ;
- $\eta(n)$  é a taxa de aprendizado;
- $\Lambda_i(x)$  é a função vizinhança centrada em volta do neurônio “vencedor”  $i(x)$ .

Sendo,  $\eta(n)$  e  $\Lambda_{i(x)}(n)$  alteradas dinamicamente durante o aprendizado em busca de melhores resultados. Deve-se ressaltar que o sucesso da formação dos *clusters* é criticamente dependente de como os parâmetros do algoritmo (taxa de aprendizado e função vizinhança) são determinados. Infelizmente não há nenhuma base teórica para a seleção desses valores, normalmente estes são determinados por um processo de tentativa e erro (HAYKIN, 1994).

### 3. Metodologia

Para o desenvolvimento do modelo, foi selecionado um único aspecto da semiótica, a codificação das informações numéricas provenientes das demonstrações contábeis em informações textuais analisadas sob a forma dos relatórios contábeis, a composição formada por essas informações que formam os relatórios anuais.

O objetivo da técnica é observar qual o grau de semiótica que existe no processo de codificação textual, uma vez que é adotada a premissa que a informação numérica das demonstrações contábeis não possui viés e representa a realidade a ser transmitida, sendo utilizada como parâmetro de mensuração.

O modelo baseado em redes neurais foi concebido partindo-se da própria analogia que o modelo neural faz ao cérebro humano. Para tanto foram observados estudos que tem como objetivo a interpretação neurobiológica da semiótica. Além disso, a Teoria da Mensuração da Informação proposta por Cropley (1998b) propõe como forma de mensuração da informação técnicas que sejam baseadas em modelos estatísticos e cibernéticos. Daí a seleção de modelo SOFM (*Self Organizing Feature Maps*), que apresenta todas as características necessárias, sendo auto-organizada, não linear, capaz de generalizar características estatísticas em seus modelos. Sendo um modelo de redes neurais constitui um dos principais campos de estudo da cibernética, portanto, validando em teoria o modelo para esta aplicação.

O modelo SOFM é um modelo auto organizado, não necessitando de exemplos em seu aprendizado, que dentre uma matriz de dados é capaz de identificar padrões comuns entre os diversos vetores de informação e dividi-los em grupos (*clusters*) que apresentem características semelhantes. Um *cluster* pode ser definido como uma estrutura cercada de propriedades, isolada externamente e coesa internamente (MANGIAMELI et al., 1996). Num exemplo prático de empresas sendo cada informação financeira uma variável, o modelo neural é capaz de dividir as empresas em grupos distintos onde cada grupo apresenta características financeiras (vetores) em comum entre seus membros e distinta dos demais. Este tipo de aplicação prático em diagnósticos financeiros é proposto por Serrano-Cinca (1996) e em grandes bases de dados por Back, Sere e Vanharanta (1998), obtendo bons resultados em ambos os casos.

A idéia do modelo de comparação é relativamente simples, utiliza-se a rede neural para detectar o padrão da informação numérica advinda das demonstrações contábeis e se obtém grupos formados por empresas com características financeiras semelhantes. Em seguida se aplica o mesmo modelo para a informação textual dos relatórios e obtém-se o mesmo número de grupos. A semiótica da relação número *vs.* texto está no padrão da codificação, ou seja, se ao codificar a informação numérica em textual foram seguidos os mesmos padrões de significado dos números pelas empresas os grupos formados pela rede neural para informação numérica e para informação textual deverão ser semelhantes. As empresas que estiverem em grupos diferentes serão as que não utilizaram o mesmo padrão semiótico entre si.

Uma aplicação semelhante à proposta foi desenvolvida por Back, Toivone, Vanharanta e Visa (2001) em artigo intitulado “*Comparing numerical data and text information from annual reports using self-organizing maps*”, entretanto a amostra era formada por empresas de diversos países que utilizam padrões contábeis (*GAAP*) diferentes, tornando impossível a comparação, uma vez que o agrupamento financeiro inicial já se encontrava comprometido.

Para eliminar esse problema no desenvolvimento do método foi utilizada a base de dados da *United States Securities and Exchange Commission*, onde podem ser obtidos os relatórios anuais completos de todas as empresas de capital aberto listadas nas bolsas de valores dos Estados Unidos e que operam nesse país, disponibilizadas no formulário *10-K* onde são dadas informações gerais sobre o negócio da empresa e sua situação financeira. Dentro destas informações obtidas pelo formulário *10-K* foram selecionadas apenas as informações textuais relacionadas à situação financeira da empresa. Para a obtenção das informações contábeis numéricas foi utilizada a base de dados Economática® uma vez que as demonstrações contábeis já se encontram padronizadas nessa base. Na amostra foram selecionadas 34 empresas para três anos consecutivos (2000, 2001, 2002), sendo estas registradas sob o código 2834 – *Pharmaceutical Preparations* segundo a classificação do *Standard Industrial Classification Code* (SIC) que divide as empresas por área de atuação. Para o desenvolvimento do modelo SOFM de classificação financeira foram utilizadas as seguintes variáveis:

<b>Balço Patrimonial</b>		<b>DRE</b>
Ativo Total	Passivo e Patrimônio Liq	Receita Líquida Operac
Ativo Circulante	Exigível Total	Custo Produtos Vendidos
Disponível e Inv CP	Passivo Circulante	Lucro Bruto
Clientes CP	Fornecedores CP	Despesas GAV e R&D
Estoques	Financiamentos CP	Outras Desp (Receit) Op
Outros Ativos CP	Outros Passivos CP	Lucro Operac Próprio
Imobilizado	Obrigações LP	Desp (Rec) Financ Liq
Intangíveis	Impostos Diferidos LP	Outras Receitas (Desp)
Outros Ativos	Outras Obrigac Diversas	LAIR
<b>DFC</b>	Patrimônio Líquido	Provisão Impost de Rend

Fluxo de Caixa Operac	Capital Preferencial	Outros
Depreciação e Amortiz	Cap Ordinário e Adicion	Lucro Op Continuadas
Out Itens do Fx Cx Oper	Lucros Acumulados	Operac Descontinuadas
Fluxo de Caixa de Inves	Other Equity	Itens Extraordinarios
Fluxo de Cx de Financm		Mudancas Princ Contab
Ef Var Tx Camb & Op Disc		Lucro Líquido
Variac Liquida de Caixa		

Tabela 1 – Variáveis Financeiras.

Para o desenvolvimento do modelo SOFM de classificação textual cada parágrafo do texto foi determinado como uma variável, dentro do limite de 25 variáveis, uma vez que acima desse nível apenas quatro empresas da amostra apresentam dados.

Para a utilização de variáveis textuais em modelos de redes neurais estas foram transformadas em vetores numéricos que viabilizaram o processamento matemático do sistema, para isso o texto foi transformado através de uma simplificação do sistema ASCII, que atribui um valor numérico a cada símbolo alfabético. Para simplificação, sinais de texto como pontos e vírgulas foram retirados do texto, assim como os espaços entre palavras, por exemplo:

*“LIQUIDITY AND FINANCIAL RESOURCES*

*A combination of cash from operations and borrowings represents the primary sources of funds to finance working capital, capital expenditures and shareholder dividends. The Company's liquidity and financial resources continued to be sufficient to meet its operating needs.”* (texto original).

*“143185383212511689316116731141922045197220*

*17415631612134164971201191941541721912134162011686419194233161020192171922021621202111217193151192520451972204995168202149316116722341913316107117321114711732111422417216832151922011682011119211414821983223821682021112741517116252014318538321251168931611673114192204519722074162131652821462205993732162121415222132120417219121316101622820”* (vetor transformado).

Após a transformação de todas as informações numéricas procedeu-se o treinamento das RNA's. Para isso foram desenvolvidas 8 redes, três para cada tipo de informação (financeira e textual), sendo dessas três uma para cada ano e outras duas (uma financeira e outra textual) com as informações dos três anos agregadas.

Assim, torna-se possível comparar a informação numérica com a informação textual ano a ano e traçar um paralelo evolutivo de cada tipo de informação no triênio 2000-2002, além da possibilidade de análise do agrupamento das informações de uma mesma empresa em anos distintos nas duas últimas redes.

Para o treinamento foram estabelecidos os seguintes parâmetros:

- SOFM Financeira e Textual(para os três anos – figura 3):
- Número de Grupos (*Clusters*): 10

- Número de Épocas de Treinamento: 1.000 + 10.000
- Taxa de Aprendizado: 0,1 e 0,01
- SOFM Financeira Geral e Textual(figura 4):
  - Número de Grupos (*Clusters*): 24
  - Número de Épocas de Treinamento: 1.000 + 10.000
  - Taxa de Aprendizado: 0,1 e 0,01

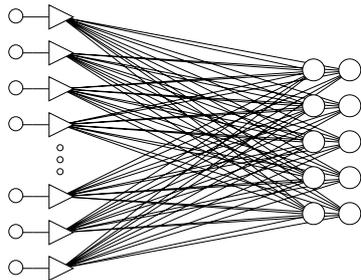


Figura 3 – Rede SOFM Anual.

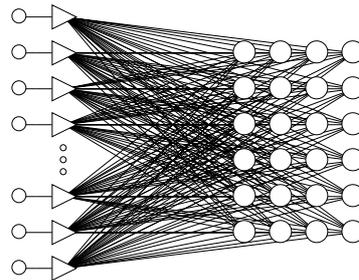


Figura 4 – Rede SOFM Geral.

A seguir são apresentados os resultados obtidos para cada rede e a comparação entre as redes de informação financeira vs. textual.

#### 4. Análise dos Resultados

Os resultados obtidos pelo agrupamento das 34 empresas podem ser analisados sob três aspectos:

- A comparação anual (2002, 2001, 2000) do grau de adequação da informação textual com relação à informação financeira, medido pelo percentual de acerto na comparação entre os grupos;
- A comparação evolutiva (2000-2002) dos grupos de mesmo tipo de informação, indicando se existe coesão entre a informação referente ao ano 2002 em relação aos mesmos dados de 2001 e 2000;
- Dentro de cada tipo de informação o agrupamento recorrente da empresa, ou seja, os dados da mesma empresa em anos distintos agrupados.

### Resultado 2002, 2001 e 2000 – SOFM Financeira vs. SOFM Textual

Grupos	Clusters 2002		Grupos	Clusters 2001		Grupos	Clusters 2000	
	Financ.	Textual		Financ.	Textual		Financ.	Textual
Grupo 1	20	11	Grupo 1	3	4	Grupo 1	3	11
Grupo 2	-	5	Grupo 2	-	-	Grupo 2	-	-
Grupo 3	6	-	Grupo 3	3	6	Grupo 3	1	5
Grupo 4	-	5	Grupo 4	-	-	Grupo 4	-	-
Grupo 5	2	1	Grupo 5	2	6	Grupo 5	3	2
Grupo 6	-	-	Grupo 6	-	-	Grupo 6	-	-
Grupo 7	3	-	Grupo 7	4	3	Grupo 7	4	1
Grupo 8	-	-	Grupo 8	-	-	Grupo 8	-	-
Grupo 9	3	12	Grupo 9	21	14	Grupo 9	20	12
<b>Acerto</b>	<b>23,53%</b>		<b>Acerto</b>	<b>42,42%</b>		<b>Acerto</b>	<b>41,94%</b>	

Tabela 2 – SOFM Financeira 2002 e SOFM Textual 2002.

Baseando-se na premissa de que a informação financeira não possui viés, no ano de 2002 apenas 23,53% das empresas foram agrupadas textualmente no *cluster* do grupo a que corresponde numericamente. Essa primeira análise indica que não existe uma padronização por parte das empresas na codificação da informação numérica em texto.

Nos casos das redes referentes aos anos de 2000 e 2001 o percentual de conformidade entre o agrupamento financeiro e textual mostrou-se significativamente maior, próximo de 42%, o que indica uma maior padronização da codificação, mas em termos gerais esse percentual é ainda muito baixo.

### Evolução 2000-2002 – SOFM Financeira e Textual

	Evolução	
		2002
<b>SOFM Financeira</b>	2001	81,82%
	2000	75,00%
<b>SOFM Textual</b>	2001	45,45%
	2000	37,50%

Tabela 3 – Evolução SOFM Financeira e Textual.

Nesse caso são comparadas as classificações, feitas pela rede SOFM Financeira, para os três anos utilizando-se como base o ano de 2002 por ser o mais recente. Pode-se observar a coesão entre as classificações feitas, denegridas apenas pelas variações financeiras que as empresas passaram nos respectivos anos, daí a diminuição do percentual de acerto conforme a maior distância do ano base, mesmo assim esse percentual pode ser considerado relativamente alto e até indicando a relação concisa do setor farmacêutico americano.

No caso da evolução do agrupamento textual o percentual de acerto com base em 2002 não é tão alto, isso possivelmente se deve ao desempenho já demonstrado em 2002 (de 23,53% apenas) na classificação textual, muito inferior aos resultados de 2001 e 2000, outro ponto relevante se observa na variação da comparação 2002-2001 e 2002-2000, pois a diferença de aproximadamente 7% tanto para a rede financeira quanto textual demonstra uma tendência de evolução semelhante entre os padrões financeiros e textuais, apesar desses apresentarem uma baixa relação entre si.

#### 4.1. Agrupamento Geral

No caso do agrupamento geral, que observa o agrupamento dos três anos das empresas, este é dividido de acordo com a informação:

Agrupamento	Clusters	
	Financeiro	Textual
<b>Acerto</b>	<b>66,67%</b>	<b>51,52%</b>

Tabela 4 – SOFM Geral (Financeira e Textual).

Nessa rede é possível observar um certo grau de agrupamento entre as mesmas empresas em anos diferentes, isso se dá pelo fato destas apresentarem alta coesão na informação, ou seja, a informação financeira do ano de 2002 é altamente dependente das informações financeiras dos anos anteriores. No caso das informações textuais este fenômeno também ocorre, indicando que estas informações seguem uma tendência semelhante a das informações numéricas. Apesar disso a tendência de agrupamento entre as informações financeiras compreende tanto agrupamento de dois como de três anos de uma empresa, enquanto que o agrupamento das empresas pelas informações textuais se dá essencialmente em conjuntos de dois anos apenas.

#### 5. Considerações finais

A contabilidade como uma linguagem é propensa a falhas na comunicação, daí a vital importância do estudo da comunicação e semiótica aplicadas a esse campo, na tentativa, se não for possível solucionar os problemas de comunicação, de ao menos entendê-los e com isso procurar evitá-los.

Para tanto são necessários estudos empíricos como este que busca de uma maneira positivista observar como se comportam os agentes na relação da percepção da informação.

Focando apenas em uma ponta do ocasional “problema” da codificação da informação financeira da empresa – essencialmente numérica – em uma análise textual.

Os resultados obtidos demonstram que os padrões de informações financeiras, aqui considerados como absolutos, não são seguidos fielmente pelas informações textuais. Isso não

significa que a informação textual não tenha seu valor ou esteja errada, mas sim que alguns aspectos inerentes à própria comunicação dificultem essa observação.

Um dos principais problemas dessa magnitude é o fato de que infelizmente existe em todas as áreas uma profusão de nomes para um único conceito e também conceitos diferentes para uma única palavra (MARTINS, 2000).

Entretanto este trabalho demonstra algumas características positivas nesse sentido. Apesar da baixa relação entre a informação financeira e textual observada (23,53% em 2002 e aproximadamente 42% em 2001 e 2000) a análise da evolução das empresas dentro dos diversos grupos indica uma tendência e uma variação proporcional muito semelhante, ou seja, os vetores de informação numérica e textual não seguem o mesmo padrão, mas historicamente (2000-2002) eles tendem a dar continuidade a informação tanto financeira quanto textual, mostrando uma alta coesão.

Estudos futuros podem demonstrar melhor essa relação, principalmente se forem adotadas melhores técnicas de pré-processamento e transformação da informação textual em vetor. Um sistema de dicionário de sinônimos no pré-processamento seria de grande valia para eliminar o problema da alta variabilidade de padrões entre as empresas e padronizar os conceitos utilizados na codificação.

Outra forma de evoluir no conhecimento sobre a relação semiótica na codificação pode ser feita alterando-se o foco de uma amostra *cross-action* para uma série temporal de uma única empresa, pois como os resultados demonstraram, existe forte coesão histórica tanto financeira como textual. Sendo que, segundo o presente trabalho demonstra, as RNA's poderiam ser capazes de realizar a codificação, ou seja, é factível o desenvolvimento de um analisador de demonstrações contábeis por meio de redes neurais desde que o mesmo seja treinado com base em um único padrão semiótico.

Portanto, as RNA's se caracterizam como uma forte ferramenta na detecção de padrões sendo de extremo auxílio no estudo da relação semiótica na contabilidade, uma vez que o modelo neural utilizado tem sua analogia no próprio sistema de compreensão semiótico humano.

## **Bibliografia**

BACK, B.; SERE, K.; VANHARANTA, H. (1998). Managing complexity in large data base using self-organizing maps. *Accounting, Management and Information Technologies*, n. 8, p. 191-210.

BACK, B.; TOIVONEN, J.; VANHARANTA, H.; VISA, A. (2001). Comparing numerical data and text information from annual reports using self-organizing maps. *International Journal of Accounting Information Systems*, n. 2, p. 249-269.

CROPLEY, D. H. (1998a). Towards formulating a semiotic theory of measurement information – Part 1. *Measurement*, n. 24, p. 237-248.

\_\_\_\_\_. (1998b). Towards formulating a semiotic theory of measurement information – Part 2. *Measurement*, n. 24, p. 249-262.

CUNNINGHAM, L. A. (2003). Semiotics, Hermeneutics and Cash: An essay on the True and Fair View. *Boston College Law School – Research Paper Series*, n. 06, p. 1-28.

- DIAS FILHO, J. M. (2001). Características Qualitativas da Informação Contábil: O Problema da Compreensibilidade à Luz da Teoria Semiótica e da Comunicação. *Dissertação de Mestrado apresentada à Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo*, São Paulo.
- DUNN, C. L.; GRABSKI, S. V. (2000). Perceived semantic expressiveness of accounting systems and task accuracy effects. *International Journal of Accounting Information Systems*, n. 1, p. 79-87.
- FAUSETT, L. (1994). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. EUA, Prentice Hall.
- IUDÍCIBUS, S.; MARTINS, E.; GELBCKE, E. R. (2000). *Manual de Contabilidade das Sociedades por Ações*. São Paulo, Editora Atlas, 5ª Edição.
- HAYKIN, S. (1994). *Neural Networks: A Comprehensive Foundation*. EUA, Prentice Hall.
- HENDRIKSEN, E. S.; VAN BREDA, M. F. (1999). *Teoria da Contabilidade*. São Paulo, Editora Atlas, 5ª Edição.
- HORNGREN, C. T.; HARRISON, W. T.; ROBINSON JR., M. A. (1996). *Accounting*. EUA, Prentice Hall, Third Edition.
- KOHONEN, T. (1989). *Self-Organization and Associative Memory*. Third Edition. USA, Springer-Verlag.
- LABOURET, V. (2001). *Semiotics and Accounting, or the Triad: Reality, Actor and Accountig*. HEC-Paris.
- LIEBENAU, J.; HARINDRANATH, G. (2002). Organizational reconciliation and its implications for organizational decision support systems: a semiotic approach. *Decision Support Systems*, n. 33, p. 389-398.
- LOPES, A. B. (2002). *A Informação Contábil e o Mercado de Capitais*. *Thonsom*, São Paulo.
- MACINTOSH, N. B.; SHEARER, T. (2000). The Accounting Profession Today: a Poststructuralist Critique. *Critical Perspectives on Accounting*, n. 11, p. 607-626.
- MANGIAMELI, P.; CHEN, S. K.; WEST, D. (1996). A Comparison of SOM neural networks and hierarchical clustering methods. *European Journal of Operational Research*, n. 93, p. 402-417.
- MARTINS, E. (2000). *Contabilidade de Custos*. 7ª Edição, Ed. Atlas, São Paulo.
- MORAES, M. B. C.; NAGANO, M. S. (2001). *Redes Neurais Artificiais em Negócios: Conceitos e Aplicações*. *1o Seminário USP de Contabilidade*.
- RUMELHART, D. E.; WIDROW, B.; LEHR M. (1994). *The Basic Ideas in Neural Network*. USA.
- SERRANO-CINCA, C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems*, n. 17, p. 227-238.
- TAFNER, M.; XEREZ, M.; RODRIGUES FILHO, I. W. (1995). *Redes Neurais Artificiais: Introdução e princípios de neurocomputação*. *Blumenau : EKO*.