

Previsão e Avaliação do Desempenho dos Contribuintes do ICMS do Estado do Ceará Utilizando as Redes Neurais Artificiais

Sérgio Ricardo Alves Sisnando

- Graduado em Ciências Econômicas (UFC)
- Graduado em Administração de Empresas (UECE)
- Especialista em Finanças (FGV/CETREDE)
- Mestre em Administração de Empresas (UNIFOR)

Marcos Airton de Sousa Freitas

- Mestre em Engenharia Civil (UFC)
- Doutorando em Engenharia (Hannover Universität – Alemanha)
- Professor da Universidade de Fortaleza (UNIFOR)

Resumo

Métodos de previsão convencionais, sejam univariados ou multivariados, têm alcançado limitado sucesso na realização de prognósticos de séries econômicas. Este fato se deve à grande dificuldade desses modelos em manipular observações decorrentes de ambientes extremamente dinâmicos, como é, por exemplo, o comportamento dos contribuintes de Imposto Sobre Circulação de Mercadorias e Prestação de Serviços (ICMS). Redes Neurais Artificiais são, em princípio, capazes de tratar com o problema de instabilidade estrutural entre as observações de uma série com parâmetros não-lineares. Nesse sentido, este trabalho procura investigar a habilidade dos modelos conexionistas em realizar avaliações e previsões acuradas do desempenho dos contribuintes do ICMS do Estado do Ceará. É proposta uma forma alternativa, por meio da identificação de padrões de comportamento. Os resultados obtidos pela aplicação de técnicas de Redes Neurais Artificiais são analisados empiricamente e confrontados com aqueles gerados pelo Modelo Atual de previsão da Secretaria da Fazenda do Estado do Ceará (Sefaz/CE).

Palavras-chave:

Redes neurais artificiais; Tomada de decisão; Métodos de previsão; ICMS.

1 – INTRODUÇÃO

1.1 – O Contexto e a Motivação da Escolha do Tema

A observação da realidade é conseqüência da manifestação de eventos complexos e incertos no decorrer do tempo. Embora, na maioria das vezes, esses eventos não sejam exatamente os mesmos – eles também não são completamente diferentes – há uma linha de continuidade, similaridade e previsibilidade que permite generalizar eventos futuros, freqüentemente de modo correto, a partir de experiências passadas.

A antecipação do comportamento futuro dos acontecimentos sempre despertou interesse nas mais diversas áreas do comportamento humano. Em finanças, por exemplo, a possibilidade de se realizar algum prognóstico permite a tomada de decisão antecipada de um investidor, em relação às expectativas do mercado. A simulação de atividades estruturalmente ligadas às empresas, considerando situações que envolvam a solução de problemas inerentes ao desenvolvimento empresarial, são subsídios importantes para o desenvolvimento de Sistemas de Informação.

Quando o conhecimento exato das leis que regem um determinado fenômeno é expresso por intermédio de equações precisas, as quais podem, a princípio, ser resolvidas, é possível prever o comportamento futuro dos eventos pela aplicação de um modelo matemático. Nesse sentido, vários procedimentos de previsão, com diferentes níveis de complexidade, foram desenvolvidos com o objetivo de solucionar esse problema. Entretanto, apesar da diversidade de métodos de previsão atualmente disponíveis – Média Móvel, Alisamento Exponencial Simples, Alisamento Exponencial Linear, Modelos Auto-regressivos, dentre outros – nem sempre a solução encontrada pelo emprego desses métodos é satisfatória. Um ponto possível dessa deficiência talvez seja a instabilidade dos parâmetros de tais métodos, mas é provável que a causa seja a ausência de não-linearidade nos modelos de previsão.

Para Souza (1989), a garantia da otimização das previsões de uma série de dados somente é alcançada adotando, como horizonte de previsão, o instante de tempo imediatamente subsequente à origem de tempo t .

A investigação do poder de previsão do modelo especificado como o mais adequado para explicar o meca-

nismo gerador das observações de uma série de dados é um processo empírico de verificação, visto que são feitas comparações entre as previsões e as observações, a fim de confirmar a habilidade do modelo matemático em descrever a estrutura definida pelos dados analisados (BARBANCHO, 1970). É importante salientar que o caráter previewal de um determinado modelo, segundo Barbancho (1970), não pode ser sustentado quando forem identificadas mudanças estruturais nas observações obtidas, a partir da origem, em relação àquelas utilizadas para a elaboração do modelo.

Outra alternativa para a realização de previsões consiste na investigação empírica de uma série de dados em busca da identificação de regularidades presentes nas observações da série de interesse. O desafio está no reconhecimento dessas regularidades, que não são sempre evidentes, sendo freqüentemente mascaradas por ruídos (REFENES; ALIPPI, 1993a). De acordo com Refenes e Alippi (1993b), há fenômenos que parecem aleatórios, sem aparente periodicidade, embora recorrentes, em um sentido geral.

A neurocomputação é uma forma alternativa de computação que procura imitar o funcionamento do cérebro humano na solução de problemas de identificação de regularidades. A identificação dessas regularidades proporcionará o reconhecimento dos padrões presentes no problema abordado, com a conseqüente observação das relações existentes entre as diversas variáveis tomadas. O interesse dos pesquisadores em uma forma de computação inspirada no funcionamento e na arquitetura do cérebro advém do fato de este último possuir características altamente desejáveis em qualquer sistema artificial, tais como: robustez e tolerância a falhas; flexibilidade; capacidade para lidar com informações probabilísticas, ruidosas ou inconsistentes; processamento paralelo; arquitetura compacta e com pouca dissipação de energia. Além disso, a capacidade de aprendizado, generalização e associação motivou o grande interesse por este tipo de computação alternativa.

A área de reconhecimento de padrões, um dos maiores objetivos das Redes Neurais Artificiais (RNA), possui uma vasta classe de problemas reais, como processamento de imagem, processamento de voz, processamento de conhecimento inexato, processamento de linguagem natural, previsão e otimização, entre outros, sendo

interessante ressaltar que muitos desses problemas ainda não possuem soluções satisfatórias dentro dos métodos convencionais existentes (PACHECO, 1991).

Devido à sua natureza indutiva, as Redes Neurais Artificiais podem se sobrepor à formulação teórica e inferir relações não-lineares complexas entre as observações de uma série de dados (REFENES; ALIPPI, 1993c). Desta forma, arquiteturas de Redes Neurais têm recebido considerável atenção, nos últimos anos, em função de suas habilidades em realizar a aprendizagem, sendo aplicadas em grande número de situações e demonstrando desempenho superior às abordagens convencionais.

Este trabalho, que é oriundo de dissertação de mestrado apresentada ao Curso de Mestrado em Administração da Universidade de Fortaleza (UNIFOR), investiga o potencial das redes neurais artificiais, na previsão e avaliação do desempenho dos contribuintes do Imposto sobre Operações Relativas à Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação (ICMS), do Estado do Ceará, dos exercícios de 1995 e 1996, cotejando os resultados alcançados com aqueles obtidos pela aplicação do modelo atual de aferição, baseado em tratamento sobre a distribuição estatística dos dados informados pelos contribuintes. Essa distribuição assemelha-se à Curva Normal, segmentando-os homoganeamente por atividade econômica. Cabe salientar que, após levantamento efetuado em outras Secretarias de Fazenda Estaduais, constatou-se a não-existência de outro modelo desenvolvido e aplicável à avaliação/previsão de desempenho de contribuintes do ICMS. Os critérios adotados, em sua maioria, caracterizam-se pela falta de sistematização e ausência de padrões.

A avaliação do desempenho dos contribuintes é a forma utilizada pelo gestor público do ICMS (Fisco Estadual) para aferir as informações econômico-fiscais declaradas, relativas às operações e prestações incluídas nas hipóteses de incidência do imposto, visando, pelos relacionamentos com outras informações do segmento de atividade econômica em que os mesmos estão inseridos e com as demais informações prestadas por outros contribuintes, identificar possíveis indícios de sonegação e, com isso, poder mensurar, através de previsões, um montante de imposto hipoteticamente sonegado pelos contribuintes.

Como o ICMS possui a característica de ser um imposto indireto e não-cumulativo, isto é, compensando-se o que for devido em cada operação relativa à circulação de mercadoria ou prestação de serviços de transporte interestadual e intermunicipal e de comunicação com o montante cobrado nas anteriores (REBOUÇAS et al., 1998), fica evidenciada a importância, para o Fisco estadual, do percentual de adicionamento empregado pelos contribuintes, possuindo uma relação direta e, na maioria das vezes, proporcional, com o valor do imposto devido. Isto é, quanto maior for a diferença entre o preço de venda/alienação e o preço de compra/aquisição das mercadorias e/ou serviços empregados, maior será a base de cálculo em que incidirá o ICMS. Fica ressaltado, dessa maneira, que o principal indicador que norteia toda a avaliação de desempenho dos contribuintes dentro da atividade econômica desempenhada por cada um é a Taxa de Adicionamento (TA) individual, significando que, quanto maior for essa taxa, melhor será o desempenho e menores serão as possibilidades de se encontrar sonegação. Por outro lado, quanto menor for essa taxa, pior será o desempenho do contribuinte e maiores serão as possibilidades de sonegação.

Avaliação e previsão do desempenho dos contribuintes do ICMS envolvem variáveis e relações não perfeitamente abrangidas pelos métodos matemáticos e estatísticos tradicionais disponíveis, devido à não-linearidade presente no relacionamento entre as diversas variáveis envolvidas. Portanto, estabelecer um método baseado nessas ferramentas implica a admissão prévia de erros e desvios que comprometerão o resultado final, provavelmente levando à tomada de decisões não-apropriadas ou incorretas, o que inviabilizaria, de antemão, a utilização de ferramentas estatísticas uni ou multivariadas.

As Redes Neurais Artificiais são inseridas nesse contexto como instrumento capaz de identificar as prováveis regularidades existentes entre as diversas variáveis que dimensionam a movimentação econômico-fiscal e caracterizam os contribuintes do ICMS do Estado do Ceará.

No ambiente de procura incessante pela otimização dos processos inerentes à modernização do Estado, insere-se a premente necessidade de se buscarem alternativas de modelos que espelhem, de maneira mais próxima à realidade dos contribuintes do ICMS, enfocando com mais acuidade aqueles que apresentem desempenho insatisfatório, pois neles certamente ha-

verá maior probabilidade de recuperação de créditos tributários em volumes superiores àqueles que apresentem desempenho satisfatório.

Os recursos da inteligência artificial (IA), em particular das redes neurais artificiais (RNA), que permitem simular a inteligência humana, podem, por meio de um mecanismo de aprendizagem, reconhecer com mais precisão e acuidade os contribuintes que necessitarão de um maior acompanhamento e fiscalização, visando sempre ao recolhimento do imposto teoricamente sonegado.

A avaliação e previsão do desempenho dos contribuintes do ICMS tornam-se prementes também em todas as secretarias de fazenda estaduais, devido à limitação de recursos e de mão-de-obra qualificada necessária à tarefa de fiscalização de todo o universo de contribuintes envolvidos. Necessariamente, há de se fazerem amostragens intencionais para seleção daqueles contribuintes que farão parte desse contexto. A questão principal surge quando temos de definir critérios e parâmetros confiáveis para elaborar tal seleção, já que o imposto é declaratório, ou seja, todas as informações sobre a circulação de mercadorias e prestação de serviços são dadas pelo próprio contribuinte, cabendo ao Fisco dois caminhos a seguir: aceitá-los como corretos e representativos ou procurar checá-los por meio de auditorias próprias, e, encontrando alguma irregularidade, cobrá-la através de lançamentos por autos de infração, acrescidas todas as penalidades cabíveis. A amostra ideal a ser selecionada seria aquela que contemplasse todos os contribuintes que apresentaram, naquele exercício em particular, o maior volume de recursos sonegados, evitando dispêndio de recursos com contribuintes que apuraram e recolhiam o ICMS corretamente. Alcançar esse ponto ótimo é o objetivo maior para uma correta racionalização dos recursos empregados.

Em um ambiente de escassez crescente de tempo e de recursos disponíveis presente em todas as organizações, sejam elas de caráter público ou privado, mais particularmente na Secretaria da Fazenda do Estado do Ceará (Sefaz), encarregada da arrecadação e fiscalização dos tributos estaduais, portanto, com a missão de prover o Estado do Ceará dos recursos necessários não somente à manutenção de toda a sua estrutura, como também de disponibilizá-los para investimentos e financiamentos visando ao desenvolvimento social da região, esse contexto não pode ter continuidade.

Este trabalho ocupa-se em cotejar as previsões oriundas do Modelo Atual empregado na SEFAZ-CE e da Rede Neural Artificial empregada. De posse das previsões geradas pelos dois modelos acima identificados, efetuam-se comparações com os resultados alcançados pelas tarefas de fiscalização *in loco* desempenhadas pelos auditores e representados pelos créditos tributários lançados por auto de infração, objetivando auferir conclusões quanto ao modelo mais apropriado, que será aquele que fornecer as previsões de crédito tributário que mais se aproximem dos valores alcançados pelas tarefas de auditoria fiscal realizadas, sendo esses considerados os valores reais de sonegação de cada contribuinte.

2 – OBJETIVOS DA PESQUISA

• Geral

Elaborar uma proposta alternativa de avaliação do desempenho dos contribuintes do ICMS do Estado do Ceará, utilizando as redes neurais artificiais (RNA), capaz de fornecer previsões mais confiáveis que aquelas apresentadas pelo modelo estatístico atualmente utilizado pela Secretaria da Fazenda do Estado do Ceará.

• Específicos

- a) Elaborar uma técnica de avaliação utilizando RNA, que permita incorporar características do processo de avaliação consideradas importantes por analistas, especialistas e tomadores de decisão;
- b) Simplificar os mecanismos de avaliação e previsão do desempenho dos contribuintes do ICMS do Estado do Ceará, de modo a permitir uma escolha mais rápida e pautada em níveis de segurança e confiabilidade superiores à escolha processada pelo Modelo atual.

3 – HIPÓTESE

A avaliação e previsão do desempenho dos contribuintes do ICMS do Estado do Ceará é um padrão que pode ser reconhecido pelas redes neurais artificiais (RNA) e os resultados encontrados são superiores àqueles gerados pelo Modelo atualmente empregado na Sefaz-CE.

Pela dinâmica empreendida pelos fatos geradores do ICMS, foram identificadas várias relações e situações

– não perfeitamente explicadas pelas limitações inerentes e pelos modelos estatísticos que tentam modelar, por equações estáticas, o comportamento e as tendências de uma série de dados. Os modelos conexionistas, em particular as Redes Neurais Artificiais (RNA), são capazes, em princípio, de identificar regularidades e relações não-lineares advindas da complexa interação de variáveis e parâmetros verificada no mundo real, tendo sensibilidade para perceber padrões na série de dados apresentada, sendo mais recomendada que o emprego de técnicas estatísticas tradicionais.

É fartamente divulgado na literatura que uma Rede Neural Artificial (RNA) pode reconhecer padrões, pois está baseada nos mesmos princípios e mecanismos da aprendizagem humana, sendo capaz de, a partir da apresentação repetida dos casos, reconhecer a existência de um “padrão” nessas apresentações, se houver, e que esse padrão pode ser captado. Ao se apresentar um “padrão” para uma RNA treinada, ela deverá procurar uma saída para o problema similar às saídas obtidas em problemas do mesmo tipo, apresentadas anteriormente.

Ao se afirmar que a avaliação e previsão do desempenho dos contribuintes do ICMS podem ser conduzidas com o uso das Redes Neurais Artificiais e que os resultados porventura encontrados serão superiores àqueles emanados do modelo atual baseado em parâmetros estatísticos, pretende-se investigar em que situações isto ocorre com segurança e, por conseqüência, como o recurso pode ser viabilizado de maneira simples e prática para esse fim.

4 – O MODELO ATUAL DE AVALIAÇÃO/ PREVISÃO DO DESEMPENHO DOS CONTRIBUINTES DO ICMS DO ESTADO DO CEARÁ

Este capítulo conceitua e explica o funcionamento do modelo atual de avaliação/previsão dos contribuintes do ICMS do Estado do Ceará, procurando explicitar a metodologia de cálculo utilizada.

Todo esse sistema está amparado em duas variáveis básicas – entradas globais e saídas globais –, que representam os valores totais informados pelos contribuintes das entradas e saídas de mercadorias e serviços ocorridas no exercício considerado, respectivamente. Primeiramente, os contribuintes são estratificados por Códigos de Atividade Econômica (vide item Dados Cadastrais do Capítulo 4); em seguida, são realizados cálculos dos índices, que serão repetidos a todos os códigos contidos no banco de dados.

O modelo atual de avaliação/previsão do desempenho dos contribuintes do ICMS do Estado do Ceará, utilizado pela Secretaria da Fazenda, foi desenvolvido e implementado, e está sistematizado atualmente em um programa elaborado em linguagem *Clipper*, em ambiente de microcomputador, denominado Sistema de Controle da Ação Fiscal (Siscaf), que efetua todas as operações e cálculos necessários, fornecendo índices individuais de cada contribuinte e índices médios do segmento de atividade econômica em que ele está inserido, em um mesmo exercício fiscal, bem como gerando uma previsão, denominada Projeção Bruta, dos valores representativos da base de cálculo da provável sonegação.

Dessa Projeção Bruta, são deduzidos todos os valores relativos às transferências ocorridas entre empresas do mesmo titular e às devoluções de compras de mercadorias, pois ambas são operações extraordinárias realizadas que não caracterizam o objeto social da empresa, e os preços envolvidos serão aqueles apurados contabilmente como custos sem a incidência de quaisquer tipos de adicionais, e as saídas de produtos tributados por Substituição Tributária, que já tiveram o imposto recolhido na fonte (em operação anterior) o que desonera de ICMS as operações realizadas posteriormente, encontrando-se a Projeção Líquida, que é a que servirá de base para a análise e avaliação do contribuinte, dentro do Código de Atividade Econômica em que está inserido.

O principal índice que norteia toda a previsão é denominado Taxa de Adicionamento (TA), que é conseguida pela razão entre a saída global, representativa dos valores das saídas de mercadorias declaradas pelo contribuinte ao longo do ano, e a entrada global, representativa dos valores das entradas de mercadorias declaradas pelo contribuinte ao longo do ano. Esse índice apresenta-nos o percentual de agregação que foi adicionado às entradas para conseguir o valor das saídas das mercadorias. Sendo essa agregação, independentemente do diferencial de alíquotas existente, a base de cálculo na qual incidirá o ICMS, já que o imposto é não-cumulativo, compensando-se o que for devido em cada operação relativa à circulação de mercadoria ou prestação de serviços de transporte interestadual e intermunicipal e de

comunicação com o montante cobrado nas anteriores por qualquer Estado, é dedutível que, quanto maior esse índice, melhor será o desempenho do contribuinte, pois maior será o montante sobre o qual incidirá o imposto.

Esse procedimento é repetido com todos os contribuintes cadastrados em um mesmo código de atividade econômica e, a partir daí, calcula-se o índice médio da taxa de adiçãoamento (TA) do setor. Como estamos tratando de uma distribuição de dados que se assemelha à curva normal, ao adicionarmos e subtrairmos da média o desvio-padrão, que avalia a dispersão dos dados em torno da média, estaremos alcançando, aproximadamente, 68,27% dos parâmetros incluídos nessa distribuição.

Para se calcular a taxa de adiçãoamento esperado (TAE) para o segmento de atividade econômica, obedecer-se-á à seguinte equação:

$$TAE = S + TA$$

Equação 3.1

Onde:

S → Desvio-padrão do setor

TA → Taxa de adiçãoamento médio do setor

A TAE fornecerá o percentual de adiçãoamento necessário para o contribuinte ser enquadrado com um desempenho muito bom dentro do seu setor de atividade no exercício sob análise.

Como forma de flexibilização e levando-se em consideração as diversas nuances que distinguem os contribuintes e as diversas conjunturas enfrentadas por cada um, utilizou-se, como parâmetro, a adoção de um fator de projeção (FP), que representaria um percentual do desvio-padrão (S), a ser adicionado à taxa de adiçãoamento médio do setor (TA), para encontrarmos uma nova TAE. Neste trabalho, adotou-se como parâmetro um fator de projeção (FP) de 50%, ou seja, para encontrar-se a TAE dos diversos segmentos envolvidos, adiciona-se à taxa de adiçãoamento médio de cada setor 50% do desvio-padrão encontrado para o mesmo. Aqui, identifica-se a avaliação de cada contribuinte por meio de um ordenamento decrescente de desempenho dentro do segmento de atividade econômica em que está inserido. Avaliar, segundo o dicionário Michaelis (MICHAELIS, 1998), representa reconhecer a grandeza, a intensidade, a força de. No caso específico, procura-se classificar o desempenho de cada contribuinte de forma individualizada, cotejando seu desempenho com o dos demais contribuintes de mesma atividade econômica, caracterizado pelos índices médios que englobam todos os componentes do estrato. Aquele contribuinte que não atingir uma taxa de adiçãoamento igual ou superior à taxa de adiçãoamento esperado (TAE), calculada para sua atividade econômica, será classificado como desempenho insatisfatório naquele exercício e entrará, como prioridade, para submissão à auditoria do ICMS. Caso o contribuinte tenha uma taxa de adiçãoamento igual ou superior à taxa de adiçãoamento esperado (TAE), calculada para seu ramo de atividade econômica, será classificado com um desempenho satisfatório e estará fora da auditoria do ICMS naquele exercício, excetuando-se dessa premissa aqueles que forem objeto de denúncias, dolos ou fraudes que envolvam o imposto, casos extras.

Para se encontrar a projeção bruta (PB), que, no presente contexto, significa a previsão, aqui entendida como predição do *quantum* de ICMS que cada contribuinte deixou de recolher, ou seja, o montante da provável sonegação, procede-se à multiplicação dessa taxa de adiçãoamento esperada (TAE), encontrada para cada setor, às entradas globais declaradas por contribuinte e deduz-se a saída global já declarada pelo mesmo. A diferença encontrada, se não houver casos de saídas a deduzir (transferências, devoluções de compras ou produtos sujeitos a substituição tributária), será o montante encontrado da provável sonegação do ICMS:

PB = (entradas globais * ((TAE/100)+1)) – saídas globais

$$PB = (entradas\ globais * ((TAE/100)+1)) - saídas\ globais$$

Equação 3.2

$$PL = PB - Saídas\ a\ Deduzir$$

Equação 3.3

Onde:

$$PL = Projeção\ Líquida$$

Saídas a Deduzir = transferências, devoluções de compras, isenções, substituição tributária.

A prioridade, para submissão à auditoria dentro de cada setor de atividade, é a ordem decrescente desses montantes de projeção bruta encontrados.

Para possibilitar melhor entendimento, coloca-se, a seguir, um exemplo numérico com valores aleatórios. A partir daí, serão calculadas a TAE (avaliação) e a PB (previsão) desse contribuinte hipotético:

- Valores Declarados pelo contribuinte:
 - Entrada Global: R\$ 220.868,29
 - Saída Global: R\$ 114.217,43
 - Transferência: zero
 - Devoluções: zero
 - Isentas: zero
 - Outras: zero
 - Índices Calculados para o CAE:
 - TA: 14,54%
 - S: 10,84%
- Cálculo da TAE, utilizando o fator de projeção (FP) – 50% para o CAE:
 - $TAE = TA + S$ (FP)
 - $TAE = 14,54 + 10,84$ (0,50)
 - $TAE = 19,96\%$
- Cálculo da projeção bruta (PB) para o contribuinte hipotético:
 - $PB = (\text{entradas globais} * ((T.A.E./100)+1)) - \text{Saídas globais}$
 - $PB = 220.868,29 * 1,1996 - 114.217,43$
 - $PB = 150.736,17$

Como inexistem valores de saídas a deduzir (transferências, devoluções, isentas, substituição tributária), o valor da projeção bruta se iguala ao da projeção líquida (PL), que será de R\$ 150.736,17, sendo esta a base de cálculo prevista de sonegação para este contribuinte, em particular.

A limitação que se impõe a este modelo é similar àquela infligida à maioria dos métodos estatísticos, ou seja, considera um número muito limitado de variáveis,

no caso, unicamente as entradas globais e as saídas globais, e desconsiderando ou tornando estáticas, no curto prazo (*ceteris paribus*), variáveis importantes, com elevada significância, para a explicação do comportamento da variável dependente. No caso específico, podem-se citar: estoques iniciais do período, estoques finais do período, crédito de ICMS, débito de ICMS, base de cálculo das entradas, base de cálculo das saídas, dentre outras elencadas no Capítulo 4, Item 4.2. Outra limitação imposta ao presente Modelo e que distorce os resultados apresentados é que os valores utilizados de entradas globais e saídas globais compreendem todas as entradas e saídas ocorridas no exercício, englobando entradas e saídas de material para uso e consumo e material do ativo permanente, operações estas que não fazem parte do objeto social da empresa, como: a compra e venda de matéria-prima, material secundário, mercadorias e produtos industrializados.

5 – REDES NEURAIS ARTIFICIAIS (R.N.A.) E O MODELO ATUAL DE AVALIAÇÃO/PREVISÃO DO DESEMPENHO DOS CONTRIBUINTES DO ICMS DO ESTADO DO CEARÁ

Neste capítulo, explicita-se a metodologia empregada na pesquisa, as limitações impostas ao presente trabalho, bem como os critérios e medidas adotados para a consecução dos resultados.

Este trabalho utilizou unicamente os dados dos contribuintes do ICMS do Estado do Ceará, pois, por meio de sondagens preliminares, não foi encontrado, no âmbito das outras Secretarias de Fazenda Estaduais, nenhum outro modelo concebido por critérios científicos que fosse destinado à previsão e avaliação de desempenho dos contribuintes do ICMS.

Os dados para composição da amostra foram extraídos do banco de dados da Secretaria da Fazenda do Estado do Ceará, onde constam informações oriundas do cadastro, ou seja, dados relativos à identificação do contribuinte, localização, objeto social, tipo de sociedade e natureza jurídica, como também dados sobre o desempenho econômico e fiscal de todos os contribuintes de ICMS do Estado, referentes aos exercícios de 1995 e 1996, informações essas que foram declaradas mensalmente pelos próprios contribuintes e são representativas de todas as operações englobadas nas hipóteses de incidência do imposto.

Como forma de torná-la operacional, excluíram-se desse universo de contribuintes aqueles enquadrados, de acordo com a legislação vigente, como microempresas ou empresas de pequeno Porte, que são estratos de contribuintes cujo faturamento bruto anual não ultrapassa a 200.000 Ufirs anuais, o que significa, aproximadamente, R\$ 212.000,00 (duzentos e doze mil reais). Isso porque, nesses estratos, encontra-se um grande número de contribuintes sem maior representatividade no total da arrecadação do ICMS, situando-se, atualmente, em níveis inferiores a 5%. Além disso, esses contribuintes inserem-se em um grupo que possui inúmeros benefícios fiscais; dentre estes estão a não-obrigatoriedade da emissão da nota fiscal na circulação de produtos ou serviços, o que inviabiliza qualquer tipo de análise de comportamento e de comparações, e a desoneração de algumas obrigações acessórias, na escrituração e apuração do ICMS.

Embora contando com banco de dados com maior volume de informações disponíveis, optou-se pela utilização dos dados nominais apresentados nos exercícios de 1995 e 1996, desprezando os dados dos anos anteriores, que estariam disfarçados pela existência de índices inflacionários que inviabilizariam a análise e exigiriam uma prévia manipulação, objetivando retirar o efeito da desvalorização da moeda sobre os números apresentados, como também dos anos posteriores, por inexistirem, até a data de realização da pesquisa, informações completas e precisas sobre os resultados dos trabalhos de auditoria realizados, impedindo qualquer comparação.

Desses contribuintes de maior significância em termos de arrecadação do ICMS, foram selecionados para compor a base de dados somente aqueles que sofreram auditoria *in loco*, nos exercícios sob análise. Pois, somente esses foram objeto de constatações e verificações das informações declaradas e os valores sonegados foram cobrados através de lançamentos por auto de infração.

5.1 – Método Proposto Para Verificação da Hipótese

Procura-se, a seguir, descrever todas as fases da pesquisa, detalhando-se mais o processo de aprendizagem da rede neural artificial, pois os valores pertinentes ao Modelo Atual já se encontram formatados através do Siscaf.

- FASE 1 – Sistematização, Classificação e Modelagem das Variáveis de Avaliação

Nesta fase, serão selecionadas, classificadas e modeladas as variáveis que servirão de base para o processo de aprendizagem da rede neural, todas retiradas de um universo de contribuintes do ICMS ativos, constantes no banco de dados da Secretaria da Fazenda do Estado do Ceará. Essa amostra intencional compreende todos os contribuintes enquadrados no Regime Normal de recolhimento que sofreram auditoria, em profundidade, no exercício de 1996. São definidas todas as variáveis de entradas e de saídas da rede neural. As variáveis de entrada são todas aquelas elencadas mais adiante no Item 5.3, compreendendo as informações provenientes do cadastro das empresas selecionadas e as informações provenientes da movimentação econômica e fiscal das mesmas. Todas essas variáveis foram modeladas para que pudessem ser utilizadas pelo aplicativo escolhido (*EASYN@*), ou seja, definindo o tipo de variável, distribuição em matriz de linhas e colunas, transformação do arquivo para o formato texto (.txt) e o número de casas decimais.

- FASE 2 – Definição da Topologia e do Algoritmo de Aprendizado da Rede Neural Artificial

- Definição quanto ao Tamanho da Rede Neural

Refenes e Alippi (1993a) dizem que o número de unidades de processamento das camadas de entrada e saída é usualmente determinado pela aplicação. No caso das camadas ocultas, a relação não é tão transparente. O ideal é utilizar o menor número possível de unidades ocultas, para que a generalização não fique prejudicada (RUMELHART; HINTON; WILLIAMS, 1986). Se o número de elementos processadores ocultos for muito grande, a rede acaba memorizando os padrões apresentados durante o treinamento. Contudo, se a arquitetura das camadas ocultas possuir unidades de processamento em número inferior ao necessário, o algoritmo *back propagation* pode não conseguir ajustar os pesos sinápticos adequadamente, impedindo a convergência para uma solução. A experiência ainda é a melhor indicação para a definição da topologia de um modelo conexionista (SURKAN; SINGLETON, 1990).

- Definição dos Pesos e Parâmetros de Aprendizado

De acordo com Freeman e Skapurad (1992), os pesos das conexões entre as camadas de uma rede neural devem ser iniciados com valores aleatórios e pequenos, para que se evite a saturação da função de ativação e a conseqüente incapacidade de realizar a aprendizagem.

À medida que o treinamento evolui, os pesos sinápticos podem passar a assumir valores maiores, forçando a operação dos elementos processadores na região onde a derivada da função de ativação é muito pequena. Como o erro retropropagado é proporcional à derivada, o processo de treinamento tende a se estabilizar, levando a uma paralisação da rede, sem que a solução tenha sido encontrada. Isto pode ser evitado pela aplicação de uma taxa de aprendizagem menor. Teoricamente, o algoritmo de aprendizagem exige que a mudança, nos pesos, seja infinitesimal (RUMELHART; HINTON; WILLIAMS, 1986). Entretanto a alteração dos pesos, nessa proporção é impraticável, pois implicaria tempo de treinamento infinito. Em vista disso, é recomendável que a taxa de aprendizado assuma maior valor, no início do treinamento e, à medida em que se observe decréscimo no erro da rede, essa taxa também seja diminuída. Diminuindo progressivamente a taxa de atualização dos pesos, o **gradiente decrescente** está apto a alcançar uma solução melhor (BEALE; JACKSON, 1990).

Outra maneira de aumentar a velocidade de convergência da rede neural artificial treinada pelo algoritmo *back propagation* é a adoção de um método chamado *momentum* (BEALE; JACKSON, 1990). O propósito desse método consiste em adicionar, quando do cálculo do valor da mudança do peso sináptico, uma fração proporcional à alteração anterior. Assim, a introdução desse termo, na equação de adaptação dos pesos, tende a aumentar a estabilidade do processo de aprendizado, favorecendo mudanças na mesma direção.

Para o caso da atualização, como se pretende utilizar o algoritmo *back propagation*, apenas duas situações são previstas: “por ciclo” e “por padrão”. Se a atualização for feita “por ciclo”, pode-se ainda definir o número de ciclos em que ocorrerá a validação. Assim, se forem definidos cem (100) ciclos para validação, esta só ocorrerá após a apresentação do centésimo exemplo. Para o caso da atualização “por padrão”, a validação apenas ocorrerá após a apresentação de todos os exemplos.

A finalidade desta FASE 2 é selecionar a topologia a ser investigada e estabelecer os índices de avaliação do desempenho dessa rede. Para este trabalho, optou-se pelas redes *Multilayer Perceptron (MLP)*, dada a sua freqüente utilização em problemas similares. O programa *EASYNN v.5.2* foi o escolhido por reunir grande quantidade de variáveis de entrada e de saída, baixo custo e

boa *interface* gráfica, não demandando equipamento específico para sua utilização.

- Método para Desenvolvimento de Aplicações de Redes Neurais Artificiais

1ª Etapa – Coleta de Dados

2ª Etapa – Separação em Conjuntos

3ª Etapa – Configuração da Rede

4ª Etapa – Treinamento

5ª Etapa – Teste

6ª Etapa – Integração

Os dois primeiros passos do processo de desenvolvimento de Redes Neurais Artificiais são a coleta de dados relativos ao problema e a sua separação em um conjunto de treinamento e um outro de testes. Essa tarefa requer uma análise cuidadosa sobre o problema para minimizar ambigüidades e erros nos dados. Além disso, os dados coletados devem ser significativos e cobrir amplamente o domínio do problema; não devem cobrir apenas as operações normais ou rotineiras, mas, também, as exceções e as condições nos limites do domínio do problema.

Normalmente, os dados coletados são separados em duas categorias: dados de treinamento, que serão utilizados para o treinamento da rede, e dados de teste, para verificar sua *performance* sob condições reais de utilização. Além dessa divisão, pode-se também usar uma subdivisão do conjunto de treinamento, criando um conjunto de validação, utilizado para verificar a eficiência da rede quanto à sua capacidade de generalização durante o treinamento, e podendo ser empregado como critério de parada do treinamento.

Determinados esses conjuntos, eles são colocados em ordem aleatória para prevenção de tendências associadas à ordem de apresentação dos dados. Além disso, pode ser necessário pré-processar esses dados, por meio de normalizações, escalonamentos e conversões de formato para torná-los mais apropriados à sua utilização na rede.

O terceiro passo é a definição da configuração da rede, que pode ser dividido em três subetapas:

- Seleção do paradigma neural apropriado à aplicação;

- Determinação da topologia da rede a ser utilizada (número de camadas, número de unidades de processamento em cada camada etc.);
- Determinação de parâmetros do algoritmo de treinamento e funções de ativação. Esse passo tem um grande impacto na *performance* do sistema resultante.

O quarto passo é o treinamento da rede. Seguindo o algoritmo de treinamento escolhido, serão ajustados os pesos das conexões. É importante considerar, nessa fase, alguns aspectos, tais como: a inicialização da rede, o modo de treinamento e o tempo de treinamento.

Uma boa escolha dos valores iniciais dos pesos da rede pode diminuir o tempo necessário para o treinamento. Normalmente, os valores iniciais dos pesos (síntese) são números aleatórios uniformemente distribuídos, em um intervalo definido. A escolha errada desses pesos pode levar a uma saturação prematura da rede.

O treinamento deve ser interrompido quando a rede apresentar uma boa capacidade de generalização e quando a taxa de erro for suficientemente pequena, ou seja, menor que um erro admissível. Assim, deve-se encontrar um ponto ótimo de parada com erro mínimo e capacidade de generalização máxima.

A quinta etapa é o teste da rede, no qual o conjunto de teste é utilizado para determinar a *performance* da rede com dados que não foram previamente utilizados. A *performance* da rede, medida nessa fase, é uma boa indicação da *performance* real.

Finalmente, com a rede treinada e avaliada, ela pode ser integrada em um sistema do ambiente operacional da aplicação. Para maior eficiência da solução, este sistema deverá conter facilidades de utilização, como uma *interface* conveniente e facilidade de aquisição dos dados.

- FASE 3 – Treinamento, Validação e Teste

Apresentam-se os procedimentos a serem executados no processo de aprendizagem da Rede Neural Artificial, objetivando conduzir a simulação, envolvendo o treinamento, validação e teste dos contribuintes selecionados. Foram seguidos os passos definidos mais adiante no Item 5.2, e estabelecidos os parâmetros de taxa de

aprendizagem, *momentum*, função de transferência, erro máximo aceitável, número de ciclos para validação.

- FASE 4 – Tabulação e Análise dos Resultados

São tabulados, apresentados, analisados e comentados os resultados obtidos pela topologia de Rede Neural definida, comparando, por meio dos indicadores estabelecidos a seguir, no Item 5.4, com os resultados apresentados pelo Modelo Atual. A finalidade da análise é a de refutar ou corroborar, com base nos resultados encontrados, a hipótese estabelecida. Ao corroborar a hipótese, vislumbra-se a transformação deste recurso computacional da Inteligência Artificial, em um procedimento simplificado, para facilitar a avaliação do desempenho dos contribuintes do ICMS, em outros exercícios fiscais.

5.2 – Modelo de Rede Neural Artificial Proposto

Entre os numerosos algoritmos de aprendizado propostos pelos métodos conexionistas, o *back propagation* é o procedimento adotado na maioria das aplicações (MASSON; WANG, 1990). A flexibilidade em manipular os mais diversos problemas, alcançando soluções satisfatórias, garantiu o sucesso desse algoritmo.

Acredita-se que o algoritmo *back propagation* seja um procedimento de aprendizagem efetivo, quando o mapeamento da entrada para a saída contenha tanto regularidades quanto exceções, sendo, em princípio, capaz de solucionar qualquer problema de classificação não-linear (HINTON, 1987). Essa característica é particularmente importante para previsões e padronizações em que muitas das tarefas consistem no tratamento de problemas de classificação não-linear.

Entretanto, ainda não está disponível qualquer método que oriente a construção de uma rede de múltiplas camadas cujo processo de aprendizagem seja realizado pelo *back propagation*. Assim, o projeto de uma rede neural é um processo exaustivo, em busca de uma topologia adequada e de valores apropriados para os parâmetros de aprendizado (REFENES; ALIPPI, 1993b).

5.3 – Arquivos de Dados Gerados

Considerando que os dados necessários a este estudo se encontravam inicialmente disponíveis apenas em plataforma de grande porte (*mainframe*), apoia-

dos em um Sistema Gerenciador de Banco de Dados e Linguagem de Programação (ADABAS/Natural2) não-adequados ao tratamento estatístico de dados, tornou-se necessário iniciar o processamento desses arquivos nesse mesmo ambiente, viabilizando a geração de um arquivo com a finalidade de armazenar os registros a serem utilizados.

Esse arquivo foi formatado de modo que pudesse ser transferido, sem maiores esforços, para plataformas menores (*PC-Windows*), onde existe maior disponibilidade de *softwares* aplicativos. Essa premissa orientou a modelagem de dados realizada e o planejamento das rotinas necessárias à geração desse banco de dados. O arquivo de dados foi transformado, para que pudesse ser trabalhado por um gerenciador de banco de dados universal (*Microsoft Access®*).

O arquivo de dados foi gerado compondo tantos registros quantos eram os contribuintes sujeitos ao regime de recolhimento normal, assim entendidos aqueles que possuísem faturamento bruto anual superior a 200.000 Ufirs, que estavam ativos nos exercícios de 1995 e 1996, e que sofreram auditoria do ICMS referente a esses períodos.

Após verificação efetuada no banco de dados, constatou-se a impossibilidade de manipulação e utilização dos dados referentes ao exercício de 1995, uma vez que alguns autos de infração, oriundos das auditorias *in loco* realizadas, que serviriam às comparações necessárias, não estavam cadastrados, distorcendo qualquer tipo de análise. A simulação da rede neural artificial contemplou, dessa maneira, apenas os valores e informações relativos ao exercício de 1996.

Dentre as variáveis disponíveis no banco de dados, que serão empregadas na modelagem e treinamento da rede neural, fazendo parte de um dos conjuntos de partições, selecionamos, empiricamente, aquelas julgadas mais significativas para medição do desempenho dos contribuintes do ICMS, separando-as em dois grandes grupos:

- **DADOS CADASTRAIS:**

1. População do Município: identifica o total da população do município recenseada, no ano de 1996;
2. Tipo do Estabelecimento: identifica o tipo de atividade desenvolvida pelo contribuinte, situan-

do-se entre as seguintes categorias: industrial, produtor agropecuário, produtor rural, prestador de serviço de transporte e de comunicação, prestador de outros serviços, comerciante atacadista, comerciante varejista e sociedade civil;

3. CAE. Principal: identifica a atividade econômica preponderante desenvolvida pelo contribuinte, classificando-a quanto à atividade, produto e público-alvo, conforme relação definida no ANEXO LVIII, do Decreto nº 24.569/97 (REBOUÇAS et al., 1998);
4. Categoria do Estabelecimento: identifica se o contribuinte é matriz ou filial;
5. Natureza Jurídica: Identifica a forma de constituição da empresa, podendo ser classificada como sociedade anônima (S/A), sociedade limitada (Ltda.), sociedade civil e empresa individual.

- **DADOS ECONÔMICO-FISCAIS:**

1. Estoque inicial do exercício: identifica o valor total dos estoques do contribuinte que estão em seu poder, de terceiros que estão em seu poder, e seus que estão em poder de terceiros, englobando matérias-primas, produtos em elaboração, produtos acabados e materiais secundários, disponíveis, no início do exercício fiscal;
2. Estoque final do exercício: identifica o valor total dos estoques do contribuinte que estão em seu poder, de terceiros que estão em seu poder, e seus que estão em poder de terceiros, englobando matérias-primas, produtos em elaboração, produtos acabados e materiais secundários, disponíveis no final do exercício fiscal;
3. Saldo credor do início do exercício: identifica o valor do saldo credor constante da conta gráfica do ICMS no início do exercício;
4. Entradas Anuais:
 - 4.1. Valor contábil das entradas anuais: identifica o valor total das entradas de produtos e serviços ocorridas durante todo o exercício;
 - 4.2. Base de cálculo das entradas anuais: é o montante total das entradas sobre o qual in-

cidirá a alíquota do imposto para encontrarmos o valor total dos créditos escriturados;

- 4.3. Imposto creditado: é o somatório de todos os valores de ICMS lançados a crédito. É a resultante da aplicação da alíquota cabível sobre a base de cálculo definida no Item 5.1;
- 4.4. Isentas/não-tributadas: é o somatório de todas as operações e prestações que entraram na empresa, amparadas pelo benefício fiscal da isenção ou que não estão alcançadas pelas hipóteses de incidência do ICMS;
- 4.5. Outras: é o somatório de todas as operações e prestações que entraram na empresa, desoneradas do imposto por força de alguma legislação específica ou que já tiveram o imposto retido, anteriormente, em alguma fase da circulação (substituição tributária).

5. Saídas Anuais:

- 5.1. Valor contábil das saídas anuais: identifica o valor total das saídas de produtos e serviços ocorridas durante todo o exercício.
- 5.2. Base de cálculo das saídas anuais: é o montante total das saídas sobre o qual incidirá a alíquota do imposto para encontrarmos o valor total dos débitos escriturados;
- 5.3. Imposto debitado: é o somatório de todos os valores de ICMS lançados a débito. É a resultante da aplicação da alíquota cabível sobre a Base de Cálculo definida no Item 5.1;
- 5.4. Isentas/não-tributadas: é o somatório de todas as operações e prestações que saíram da empresa amparadas pelo benefício fiscal da isenção ou que não estão alcançadas pelas hipóteses de incidência do ICMS;
- 5.5. Outras: é o somatório de todas as operações e prestações que saíram da empresa desoneradas do imposto por força de alguma legislação específica ou que já tiveram o imposto retido anteriormente, em alguma fase da circulação (substituição tributária).

6. Valor Anual da Arrecadação: é a resultante de todos os valores arrecadados pelo contribuinte durante o exercício fiscal. Está estratificado por códigos de quatro algarismos que indicam o tipo de receita do ICMS arrecadada, podendo ser: ICMS-normal, ICMS-importação, ICMS-substituição entradas, ICMS-substituição saídas e ICMS-antecipado. indicam também o tipo de atividade geradora do imposto: comércio; indústria; agricultura; pecuária; aves; piscicultura; serviços de comunicações; transportes, combustíveis líquidos, gasosos e lubrificantes; energia elétrica, minerais e autos de infração;
7. GIM: identifica o número de meses do exercício em que cada contribuinte apresentou movimentação econômico-fiscal.

Todos as variáveis envolvidas, nos Itens 4 e 5, estão estratificadas por código fiscal da operação (CFOP), que possui três algarismos e destina-se a aglutinar, em grupos homogêneos, nos livros fiscais, nas guias de informação e em todas as análises de dados, as operações e prestações realizadas pelos contribuintes do ICMS, conforme estabelecido no Anexo LVII do Decreto nº 24.569/97 (REBOUÇAS et al., 1998), tendo a seguinte representação genérica:

- Iniciando com 1 – Operações e prestações de entradas provenientes do mesmo Estado;
- Iniciando com 2 – Operações e prestações de entradas provenientes de outros Estados da federação;
- Iniciando com 3 – Operações e prestações de entradas provenientes do exterior (importação);
- Iniciando com 5 – Operações e prestações de saídas destinadas ao mesmo Estado;
- Iniciando com 6 – Operações e prestações de saídas destinadas a outros Estados da federação; e
- Iniciando com 7 – Operações e prestações de saídas destinadas ao exterior (exportação).

A partir dessa estratificação, cada variável dos itens 4 e 5 foi subdividida em 18 outras variáveis. Todas as

classificações obedeceram à codificação contida no ANEXO LVII do Decreto 24.569/97 (CEARÁ, 1997), assim consideradas:

Desse entendimento, verifica-se que o Banco de Dados que será adotado para a Rede Neural Artificial será uma matriz composta de 2.656 linhas (representando os contribuintes selecionados e que estão discriminados no Quadro 2 a seguir) e 190 colunas (5 representando os dados cadastrais, 180 representando os itens 4 e 5 dos dados econômico-fiscais e 5 representando os itens 1, 2, 3, 6 e 7 dos dados econômico-fiscais). Essas 190 colunas são identificadoras das variáveis de entrada na Rede Neural, devendo existir, na camada de entrada, tantos neurônios quantos forem as variáveis de entrada. Pela análise do Quadro 2, constata-se que quase 78% dos contribuintes selecionados participam da atividade de comércio (varejista e atacadista), representando a preponderância dessa atividade, no âmbito dos contribuintes do ICMS.

Para definição dos subconjuntos de treinamento, validação e teste da rede neural, utilizamos os padrões e

regras estabelecidos por Pretchelt (1994). Todos os subconjuntos foram selecionados de forma aleatória e excludente, não havendo repetição de exemplos em outros subconjuntos. Com isso, pretende-se evitar que a rede fique viciada, o que poderia comprometer a sua capacidade de generalização do contexto apresentado.

Para comparação dos resultados encontrados com os dois modelos adotados no presente trabalho, adotam-se como parâmetro os valores lançados através de autos de infração provenientes dos trabalhos de auditoria, em profundidade, realizados nos contribuintes (variável de saída da rede neural artificial). Neste tipo de auditoria, são conferidas e analisadas todas as operações, em um determinado exercício fiscal, que tiveram repercussão ou influência na apuração do ICMS, abrangendo, além da auditoria fiscal propriamente dita, a auditoria contábil e a auditoria de produção/estoques, o que significa uma verificação completa das informações econômico-fiscais do contribuinte. Qualquer diferença encontrada é cobrada através de lançamento do crédito tributário pela lavratura do competente auto de infração.

Cod	CFOP ENGLOBADOS
1.11	1.11 a 1.14; 5.11 a 5.17
1.21	1.21 a 1.14; 5.21 a 5.26
1.31	1.31 a 1.34; 5.31 a 5.34
1.41	1.41 a 1.44; 5.41 a 5.45
1.51	1.51 a 1.55; 5.51 a 5.53
1.61	1.61 a 1.65; 5.61 a 5.79
1.91	1.91 a 1.99; 5.91 a 5.99
2.11	2.11 a 2.14; 6.11 a 6.19
2.21	2.21 a 2.24; 6.21 a 6.26
2.31	2.31 a 2.34; 6.31 a 6.34
2.41	2.41 a 2.44; 6.41 a 6.45
2.51	2.51 a 2.55; 6.51 a 6.53
2.61	2.61 a 2.65; 6.61 a 6.79
2.91	2.91 a 2.99; 6.91 a 6.99
3.11	3.11 a 3.13; 7.11 a 7.17
3.21	3.21 a 3.24; 7.31 a 7.34
3.41	3.41 a 3.44; 7.41 a 7.61
3.91	3.91 a 3.99; 7.99

Quadro 1 – Consolidação das CFOPs

Fonte: Ceará (1997).

Nessa altura, por envolver um número muito grande de contribuintes de segmentos de atividades econômicas muito diversos uns dos outros, alguns com menor número de empresas e maior representatividade na arrecadação do ICMS, outros com maior contingente de empresas, porém sem grande participação no total da arrecadação do imposto, como também pela enorme variedade de códigos de atividades econômicas contemplados, optou-se pela aglutinação em um único banco de dados representativo dos valores do exercício de 1996. Neste ínterim, a variável até então discriminante no Modelo atual, código de atividade econômica (CAE) passou a ser mais uma variável introduzida no processo de aprendizagem da rede neural.

O Quadro 2, mostra a estratificação do universo de contribuintes contemplados na pesquisa pelo código de atividade individual predominante:

Por impedimento legal, Artigo 198 do Código Tributário Nacional (BRASIL, 1989), “é vedada a divulgação, para qualquer fim, por parte da Fazenda Pública ou de seus funcionários, de qualquer informação, obtida em razão de ofício, sobre a situação econômica ou financeira dos sujeitos passivos ou de terceiros e sobre a natureza e o estado dos seus negócios ou atividades”. Dessa forma, a identificação dos contribuintes foi omitida do Banco de Dados, sendo substituída por números-índice.

Como ferramentas auxiliares, utilizou-se o *software easynn v.5.2* e os aplicativos desenvolvidos em linguagem *Clipper*, denominados Sistema de Controle da Ação Fiscal – Siscaf e que são atualmente empregados na Sefaz-CE. Todos eles foram instalados em um microcomputador Pentium® II Celeron 300 Mhz, que foi o *hardware*-padrão para o desenvolvimento desta pesquisa.

GRUPO	DISCRIMINAÇÃO	NÚMERO DE CONTRIBUINTES
10	Indústria de Produtos de Minerais Não-metálicos	38
11	Indústria Metalúrgica	28
12	Indústria Mecânica	07
13	Indústria de Material Elétrico, Eletrônico e de Comunicações	04
14	Indústria de Material de Transporte	07
15	Indústria de Madeira	15
16	Indústria de Mobiliário	19
17	Indústria de Papel e Papelão	03
18	Indústria de Borracha	04
19	Indústria de Couro, Pele e Produto Similar	04
20	Indústria Química	27
21	Indústria de Produto Farmacêutico e Veterinário	02
22	Indústria de Perfumaria, Sabões e Velas	12
23	Indústria de Material Plástico	16
24	Indústria Têxtil	21
25	Indústria de Vestuário, Calçado, Artefato de Tecido, Couro e Pêlo	104
26	Indústria de Produto Alimentar	52
27	Indústria de Bebida	16
29	Indústria Editorial e Gráfica	07
30	Indústria Diversa	08
31	Indústria e/ou Serviço de Utilidade Pública	03
32	Indústria e/ou Serviço de Construção	04
40	Agropecuária	68
50	Serviços de Transporte	72
53	Serviços de Reparação, Manutenção e Conservação	02
57	Serviços de Escritório de Gerência e Administração e Depósito Fechado	02
58	Serviços de Construção Civil	05
60	Comércio Atacadista	549
61	Comércio Varejista	1.557
TOTAL		2.656

Quadro 2 – Estratificação do universo de contribuintes

Fonte: Arquivo de dados trabalhados.

5.4 – Medidas de Avaliação dos Resultados

Com o objetivo de avaliar os resultados das RNAs e do Modelo Atual, em vigor na SEFAZ-CE, de forma consistente, foi realizada a aferição tomando-se como base a *Equação 4.1*:

$$I = \frac{\sum |SG_{atual} - SG_{real}|}{\sum |SG_{rede} - SG_{real}|}$$

Equação 4.1

Onde:

SG_{atual} – Saídas globais previstas pelo modelo atual;

SG_{real} – Saídas globais reais;

SG_{rede} – Saídas globais previstas pela rede neural.

Nesta *Equação* serão considerados apenas os valores modulares, representativos dos números absolutos encontrados para cada contribuinte.

Por intermédio do indicador (I) pode-se concluir que, se:

I > 1 – Modelo da rede neural apresentou melhor desempenho;

I < 1 – Modelo atual apresentou melhor desempenho;

I = 1 – Ambos os modelos apresentaram desempenhos semelhantes. Neste caso, complementa-se a análise utilizando o desvio-padrão como medida do grau de dispersão das diferenças. O modelo que apresentar melhor desempenho será aquele que obtiver o menor desvio-padrão.

O indicador (I) apenas informará qual o modelo que apresentou melhores resultados, porém, para auferir-se o percentual de superioridade de um em relação ao outro, utilizar-se-ão os indicadores I_{ATUAL}, que representarão o resultado apresentado pelo modelo atual, e I_{REDE}, que representará o resultado apresentado pelo modelo definido pela rede neural artificial. Assim, tem-se:

$$I_{ATUAL} = \left(\frac{\sum SG_{atual}}{\sum SG_{real}} \right) \times 100$$

Equação 4.2

$$I_{REDE} = \left(\frac{\sum SG_{rede}}{\sum SG_{real}} \right) \times 100$$

Equação 4.3

Por meio desses indicadores, *Equações 4.2 e 4.3*, pode-se identificar qual Modelo, em forma percentual, mais se aproxima dos valores definidos como saídas globais reais. O indicador que mais se aproximar de 100% (cem por cento) representará o modelo mais apropriado para análise do problema em questão.

6 – APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Neste capítulo, são apresentados e analisados os resultados dos experimentos realizados. Primeiramente, enfocam-se os parâmetros adotados e os resultados alcançados com a rede neural artificial utilizada. Em seguida, calculam-se os indicadores para identificar o modelo mais apropriado para o contexto sob análise, visando corroborar ou não a hipótese levantada.

6.1 – Apresentação da Rede Neural Artificial

Cumpridas as etapas definidas na FASE I do item 5.1, em que foram selecionadas as variáveis que serviram de parâmetros de entrada e de saída para a RNA, classificadas e modeladas para que pudessem ser utilizadas pelo *software*-padrão utilizado para o desenvolvimento deste trabalho, passou-se para a FASE 2, em que seria necessária a definição da topologia e do algoritmo de aprendizado da RNA.

Quanto à definição do algoritmo de aprendizagem, assunto já abordado no item 5.2, acredita-se que não haja outras considerações a serem feitas. No tocante à topologia da rede, utilizou-se, inicialmente, uma rede com o formato 190-6-1, ou seja, 190 neurônios na camada de entrada, 6 neurônios na camada oculta ou escondida e 1 neurônio na camada de saída, sem agrupamentos em subconjuntos, com arquitetura totalmente conectada.

Seguindo-se a FASE 3 e de acordo com o padrão de Pretchelt (1994) proposto no método de verificação, particionamos a amostra em três subconjuntos, assim definidos: 50% (1.327 exemplos) para treinamento, 25% (665 exemplos) para validação e 25% (664 exemplos) para teste e previsão. Os dados de teste e previsão foram es-

colhidos de forma intencional pela seleção dos primeiros 664 casos do Banco de Dados. Os dados de validação e treinamento foram escolhidos de forma aleatória pelo próprio aplicativo. Procurou-se, desta forma, evitar que os dados apresentados de forma seqüenciada tornassem a rede viciada em um certo tipo de apresentação, o que dificultaria o poder de generalização para outros contextos não submetidos à RNA.

Os pesos iniciais, representativos do processo de síntese, foram definidos de forma randômica pelo próprio *software*. Foram mantidas constantes as taxas de aprendizagem (η) e momento (α), respectivamente em 0,6 e 0,8, parâmetros esses sugeridos pelo próprio aplicativo, após definida a topologia da RNA. O erro médio máximo foi estipulado em 2%. Para validação, utilizou-se, como critério de parada da RNA, um percentual de 95% do subconjunto, podendo existir uma variação de até 48% da taxa de erro médio máximo definida. Foi adotado o critério de validação “por ciclo”, sendo efetivada a cada 5 ciclos.

Com esses parâmetros, a RNA atingiu um erro médio máximo de 1,9871% em 166.459 ciclos de treinamento e validação.

Na FASE 4, passou-se a apresentar e analisar os resultados encontrados pela RNA, comparando-se os resultados encontrados com aqueles gerados pelo Modelo Atual empregado no âmbito da Secretaria da Fazenda do Estado do Ceará.

6.2 – Análise e Comparação dos Resultados

Aplicando-se os indicadores preestabelecidos no Item 5.4, encontramos os seguintes resultados:

I - Subconjunto de Teste:

Neste agrupamento, consideraram-se os contribuintes de números 01 a 925, apresentados de forma seqüencial, no banco de dados, e encontraram-se os seguintes resultados:

$$I = 1,4918$$

$$I_{REDE} = 111,10\%$$

$$I_{ATUAL} = 6,05\%$$

Conclui-se que o modelo de rede neural apresentou resultado bastante superior àquele oriundo do mode-

lo atual, corroborando, plenamente, a hipótese retro-apresentada. Da análise dos indicadores I_{REDE} e I_{ATUAL} , pode-se deduzir, também, que as previsões geradas pela RNA foram melhores que as produzidas pelo Modelo Atual, pois o I_{REDE} ajustou-se em 111,10%, enquanto o I_{ATUAL} foi de 6,05%.

Pela observação do Gráfico 1, a seguir, em que consta um comparativo entre os valores reais e os gerados pela RNA treinada e pelo Modelo Atual, pode-se claramente constatar que os valores obtidos através do modelo conexionista tiveram uma aproximação maior com os valores reais que o Modelo Atual, gerando previsões mais fidedignas com a realidade. Neste gráfico, foram considerados unicamente os dados contidos no subconjunto de teste. No eixo das abscissas foram colocados os números-índice representativos dos contribuintes deste subconjunto, enquanto que, no eixo das ordenadas, estão contemplados os valores em reais (R\$) da variável de saída.

II) Subconjuntos de Treinamento e Validação:

Nesses agrupamentos foram considerados os contribuintes de números 926 a 3.609, encontrando-se os seguintes resultados:

$$I = 1,9983$$

$$I_{REDE} = 102,79\%$$

$$I_{ATUAL} = 4,72\%$$

Como seria lógico, os subconjuntos de treinamento e validação apresentaram resultados superiores àqueles do subconjunto de teste, encontrando-se índices e previsões mais acurados e mais próximos do real, contudo, corroborando também a hipótese levantada. Os indicadores I_{REDE} e I_{ATUAL} demonstram essa situação. Enquanto o I_{REDE} suplantou em apenas 2,79% o real, o I_{ATUAL} situou-se no patamar de 95,28% inferior ao real, demonstrando-se inequivocamente a hipótese apresentada.

6.3 – Considerações Finais

A seleção de uma rede não implica que a escolhida seja a única. Deixou-se de contemplar um grande número de redes com outros tipos de topologias. Procurou-se, acima de tudo, garantir que o processo de aprendizagem não se constituísse em uma forma de “decorar padrões de previsão”. Neste trabalho, nenhum fenômeno de *overfitting* ou *underfitting* foi observado.

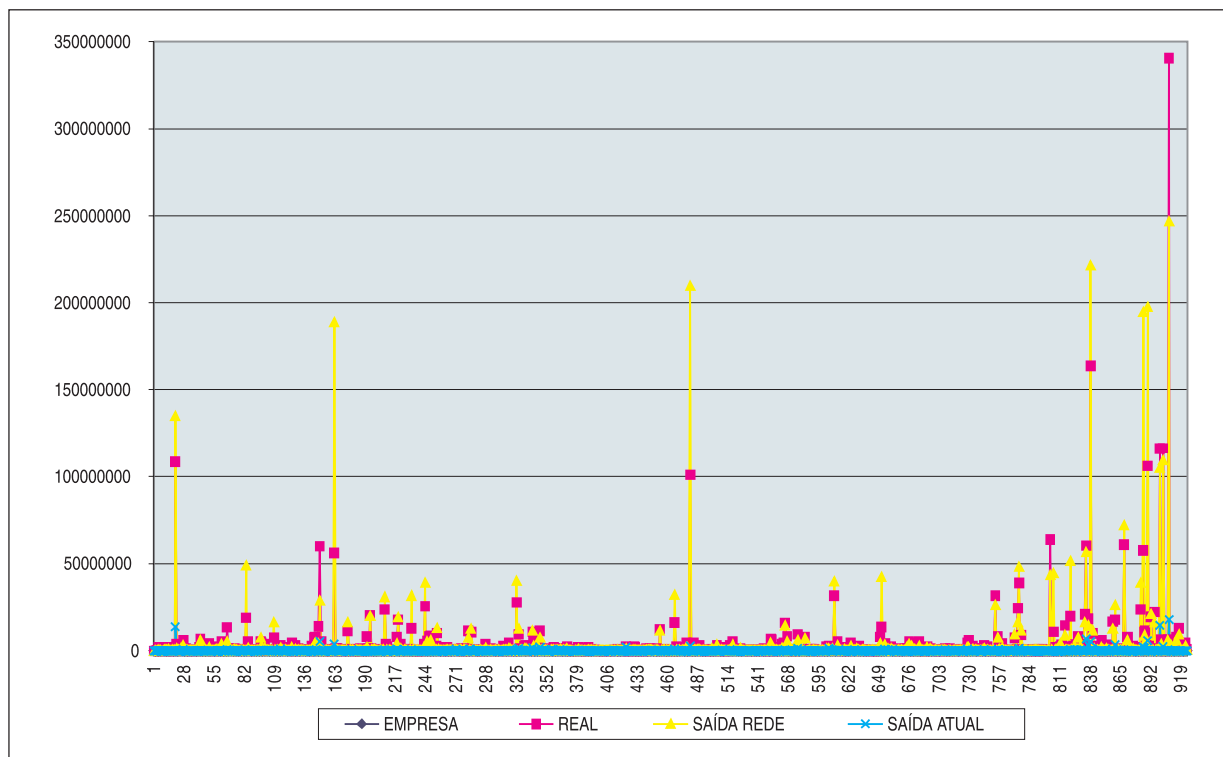


Gráfico 1 – Comparativo entre a RNA e o modelo atual

Fonte: Elaboração dos autores.

É importante frisar que não se deve tentar buscar, no sistema de ponderação de uma RNA, a leitura dos pesos das variáveis para a readequação a uma ferramenta ou técnica alheia à que foi testada. Salienta-se que o referido sistema é o resultado de um processo de aprendizagem em que a rede consegue “extrair” características significativas dos exemplos apresentados. A partir daí, é possível aplicá-las a outros casos, de acordo com os dados apresentados para previsão.

Abstract

Conventional methods of forecast, are univariéd or multivariéd, have reached limited success in the accomplishment of prognostics of economic series. This behavior is due to the great difficulty of these models in manipulating decurrent extremely dynamic environment comments, as, for example, the behavior of the contributors of Tax On Circulation of Goods and Installment of Services (ICMS). Artificial Neural nets are, firstly, capable to deal with the problem of structural instability between the comments of a series with nonlinear parameters. In this direction this work looks for to investigate the ability of the connection models in evaluations and acurated forecasts of the performance of the contributors of the ICMS of the

State of the Ceará. An alternative form is proposed, by means of the identification of behavior standards. The results gotten with the application of techniques of Artificial Neural Nets are empirically analyzed and faced with those generated by the Current Model of forecast of the Secretaria da Fazenda of the State of Ceará (SEFAZ/CE).

Key words:

Artificial neural nets, taking of decision, forecast

REFERÊNCIAS

- BARBANCHO, A. G. **Fundamentos e possibilidades da econometria**. Rio de Janeiro: Fórum Editora, 1970.
- BEALE, R.; JACKSON, T. **Neural computing: an introduction**. Bristol: Adam Hilger, 1990..
- BRASIL. **Código tributário nacional**. Lei 5.172. 18. ed. São Paulo: Saraiva, 1989.
- CARVALHO, A. C. P. L. F. et al. **Fundamentos de redes neurais artificiais**. Rio de Janeiro: DCC/IM, COPPE/Sistemas, NCE/UFRJ, 1998.

CEARÁ. Decreto 24.569/97 de 31 de julho de 1997. Regulamenta o ICMS no Estado do Ceará e dá outras providências. **Diário Oficial do Estado do Ceará**, Fortaleza, 31 jul. 1997.

DAMASCENO, A. P. **Um modelo multivariado para a avaliação dos contribuintes que requerem parcelamento de débitos**. 1997. Dissertação (Mestrado em Administração de Empresas) – Universidade de Fortaleza- UNIFOR, Fortaleza, 1997.

EASYN@ - Steve Wolstenholme. **Neural network software for windows**. Disponível em: <<http://www.tropheus.demon.co.uk>>. Acesso em: jul. 2000.

FACHIN, O. **Fundamentos de metodologia**. São Paulo: Atlas. 1993.

FARNUM, N. R.; STANTON, L. W. **Quantitative forecasting methods**. [S.l.]: PWS-KENT Publishing Company, 1989.

FEIGENBAUM E.; BARR, A. **The handbook of artificial intelligence**. [S.l.]: William Kaufmann Inc., 1981. V. 1.

FREEMAN, J.; SKAPURAD. **Neural network algorithms, applications and programming techniques**. [S.L.]: Addelfon-Wefley, 1992

FURTADO, A. N. D. **Uma nova abordagem na avaliação de projetos de transporte**: o uso das redes neurais artificiais como técnica para avaliar e ordenar alternativas. 1998. Tese (Doutorado Engenharia de Transportes) – Escola de Engenharia de São Carlos, Universidade de São Paulo-USP, São Paulo, 1998.

GRANGER, C. W. J.; NEWBOLD, P. **Forecasting economic time series**. New York: Academic Press, 1977.

HINTON, G. **Connectionist learning procedures**. [S.l.]: Carnegie-Melon University, 1987.

HOEL, P. G. **Estatística elementar**. São Paulo: Atlas. 1980.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4th. ed. New Jersey: Prentice-Hall, 1998.

KASHIGAN, S. K. **Multivariate statistical analysis**: a conceptual introduction. 2nd ed. New York: Radius Press, 1991.

LE CUN, Y. **Generalisation and network design strategies** Toronto: University of Toronto, 1989.

MASSON, E.; WANG Y. **Introduction to Computation and Learning in Artificial Neural Networks**. European Journal of Operation Research, North-Holand, 1990.

MENDENHALL, W. *et al.* **Statistic for management and economics**. 7. ed. [S.l.]: Wadsworth Inc., 1993.

MICHAELIS: moderno dicionário da língua portuguesa. São Paulo: Companhia Melhoramentos, 1998. p. 268.

MILLS, F. C. **Métodos estadísticos aplicados a la economía y a los negocios**. Madrid: Aguillar Ediciones, 1967.

MILONE, G.; ANGELINI, F. **Estatística aplicada**. São Paulo: Atlas, 1995.

MORRISON, D. F. **Multivariate statistical methods**. 3. ed. [S.l.]: McGraw-Hill, 1990.

PACHECO, M. A. C. **A neural-risc processor and parallel architecture for neural networks**. 1991. Tese (PhD em Ciência da Computação) - Department of Computer Science UCL, Londres, 1991.

PRETCHELT, L. **PROBEN 1**: a set of neural network benchmark problems and benchmarking rules. [S.l.]: Fakultät für Informatik, 1994. (Technical Report 21/94).

REBOUÇAS, O. *et al.* **Regulamento do ICMS do estado do Ceará**. Ed. anot. e atual. Fortaleza: Expressão Gráfica e Editora, 1998.

REFENES, A. N.; ALIPPI, C. Currency exchange rate prediction and neural networks design strategies. **Neural Computing and Applications Journal**, London, v. 1, n. 1, 1993a.

REFENES, A. N.; ALIPPI, C. **Financial modelling using neural networks**. London: University College London, 1993b.

REFENES, A. N.; ALIPPI, C. **Stock performance modeling using neural networks**: a comparative study with regression models. London: University College London, 1993c.

RUDIO, F. V. **Introdução ao projeto de pesquisa científica**. 10. ed. Petrópolis: Vozes. 1985.

RUMELHART, D; HINTON, G.E.; WILLIAMS, R.J. Learning internal representations by error propagation. In: _____. **Parallel distributed processing**: exploration in the microstructure of cognition. Cambridge : M.I.T. Press, 1986.

RUMELHART, D.; McCLELLAND, J. **Parallel distributed processing explorations in the microstructure of cognition**. Massachusetts: The M.I.T. Press. Cambridge, 1986.

RUMELHART, D.; ZIPSER, D. **Discovering by competitive learning**. Massachusetts: The M.I.T. Press. Cambridge, 1986.

SPIEGEL, M. R. **Schaum's outline of theory and problems of statistics**. [S.l.]: MacGraw Hill, 1971.

STEVENS, J. K. **Reverse engineering the brain**. [S.l.]: BYTE, 1985.

SCHALKOFF, R. **Artificial intelligence**: an engineering approach. São Paulo: McGraw-Hill, 1990.

SOUZA, R. C. Modelos estruturais para previsão de séries temporais: abordagens clássica e baynesiana. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 17., Rio de Janeiro, 1989. **Anais...** Rio de Janeiro: IMPA, 1989.

SURKAN, A.; SINGLETON, C. Neural networks for bond rating improved by multiple hidden layers. In.: International Joint Conference on Neural Networks, 1990, San Diego. **Proceedings...** San Diego: IEEE, 1990.

TAFNER, M. A. et al. **Introdução e princípios de neurocomputação**: redes neurais artificiais. Brasília, DF: EKO, 1998.

WASSERMAN, P. D. **Neural computing**: theory and practice. New York: Van Nostrand Reinhold, 1989.

WHEELWRIGHT, S. C.; MAKRIDAKIS, S. **Forecasting methods for management**. 4th ed. New York: John Wiley & Sons, 1985.

WINSTON. **Artificial intelligence**. 2nd Edition, Massachusetts: Addison-Wesley, 1984.

Recebido para publicação em 22.07.2002.