
Previsão de vendas no varejo por meio de redes neurais

Fernando Carvalho de Almeida
Antônio Fabrizio Lima Passari

RESUMO

Neste trabalho, explora-se o uso de redes neurais na previsão de vendas no varejo. Com essa técnica foram elaboradas previsões a partir de dados históricos de vendas de produtos de uma empresa do varejo. Foram gerados modelos com o uso de dados de venda de cada um dos produtos da base de dados, semana a semana. Os resultados sugerem que a modelagem por meio de redes neurais artificiais pode ser considerada adequada para a previsão de demanda de produtos no nível individual (produto a produto). Os modelos obtidos com o uso da metodologia proposta podem prever as vendas de produtos no curto prazo com maior precisão do que as técnicas *naïve* não-ajustada e de regressão linear, mais frequentemente utilizadas.

Palavras-chave: previsão de vendas, varejo, redes neurais.

1. INTRODUÇÃO

O avanço da capacidade de processamento dos computadores nos últimos anos deu grande impulso ao uso de métodos quantitativos em Administração, fazendo crescer nas empresas o número de suas aplicações (BERRY e LINOFF, 1997; GARGANO e RAGGAD, 1999).

Técnicas comumente chamadas de mineração de dados, ou *data mining*, utilizam-se fortemente de ferramentas de tratamento de dados, técnicas estatísticas e outras abordagens quantitativas para a extração de conhecimento contido nas bases de dados acumulados por uma empresa. Esses dados são acumulados à medida que a empresa realiza suas atividades e processos de negócios e registra o histórico dessas atividades geradas tanto internamente quanto no relacionamento com clientes e fornecedores. Por meio das técnicas de *data mining*, os dados históricos são explorados para detectar padrões ou fazer previsões a partir da experiência contida nos dados.

As técnicas de mineração de dados incluem não só métodos de inteligência artificial (sistemas especialistas, lógica *fuzzy*, entre outros), árvores de decisão, métodos de indução de regras, redes neurais, entre outras (GARGANO e

Recebido em 22/setembro/2004
Aprovado em 02/dezembro/2005

Os autores agradecem a colaboração da IBM Brasil e do Sr. Artur Miranda por ceder a eles o *software Intelligent Miner 6.1*, que permitiu a realização da pesquisa relatada neste artigo.

Fernando Carvalho de Almeida é Professor Doutor do Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (CEP 05508-010 — São Paulo/SP, Brasil).

E-mail: fcalmeida@usp.br

Endereço:

Universidade de São Paulo

FEA — Departamento de Administração

Avenida Professor Luciano Gualberto, 908 — FEA 1

— Sala G-122

05508-010 — São Paulo — SP

Antônio Fabrizio Lima Passari é Mestre em Administração de Empresas pela Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (CEP 05508-010 — São Paulo/SP, Brasil).

E-mail: parqtec@uol.com.br

RAGGAD, 1999), como também o uso de métodos estatísticos convencionais (HAND, 1998). Buscando através dessas técnicas novas informações e conhecimento, procura-se, em certos casos, **aprender** padrões, correlações e interdependências a partir de relacionamentos escondidos nos dados.

Neste trabalho, explora-se o uso dessas técnicas para a criação de modelos voltados à detecção de padrões e à previsão de demanda no varejo.

O uso de técnicas quantitativas na previsão de vendas pode ser encontrado em diversos estudos. Tang, Almeida e Fishwick (1991), Luxhoj, Riis e Stensballe (1996), Ansuji *et al.* (1996), Kuo e Xue (1998) e Spedding e Chan (2000) pesquisaram a adoção de modelos de previsão de vendas agregadas, isto é, para toda uma corporação ou loja.

Outros autores exploraram a previsão de venda de produtos de maneira individualizada, a partir de séries históricas das vendas de cada produto (BARASH e MITCHELL, 1998; GORDON, 1998; ALIBAIG e LILLY, 1999). Esses estudos utilizaram técnicas de modelagem de séries temporais, tomando como entrada dos modelos os valores históricos das vendas de cada produto no tempo, e construíram um modelo distinto para a previsão de vendas de cada produto. Mais recentemente, Frank *et al.* (2003) exploraram o uso de redes neurais na previsão de vendas de roupas para mulheres.

Outros autores estudaram o impacto de diversas variáveis na previsão de demanda de um produto isolado, chamada previsão causal (MENTZER e KENT, 1999; SUBRAHMANYAN, 2000).

Poucos autores, contudo, exploraram modelos que incorporassem a interação entre os produtos e a influência cruzada em suas vendas. Alguns autores, como Ben-Akiva e Gershensfeld (1998) e Guadagni e Little (1998), investigaram o relacionamento entre produtos na previsão de vendas, mas esses estudos pouco utilizaram dados históricos, baseando-se, principalmente, em análises exploratórias e pesquisas de grupo. Vê-se, então, uma lacuna na bibliografia sobre previsão, que é o estudo do relacionamento e influência entre os produtos na previsão de vendas que incorpora informações acerca de mais de um produto simultaneamente.

Neste estudo, explora-se a previsão de vendas de um produto (demanda individual), levando em consideração o impacto da demanda simultânea (ou das ações sobre a demanda) de demais produtos relacionados às vendas desse produto. Avaliam-se os relacionamentos existentes entre os produtos, utilizando para tanto a técnica de redes neurais.

2. PREVISÃO DE VENDAS

Para elaboração e teste dos modelos de previsão, três aspectos básicos foram abordados na pesquisa aqui relatada:

- nível de agregação dos dados;
- seleção de variáveis preditivas e o composto de *marketing*;
- seleção de produtos afins — identificação de centro de interesse.

Esses três aspectos são explanados a seguir.

2.1. Nível de agregação dos dados

As abordagens empregadas para a previsão de vendas (ROBERTS, 1998) variam consideravelmente segundo o nível de agregação dos dados (individual ou agregado) e segundo o tempo de existência do produto no mercado (produto novo ou produto já no mercado), conforme o quadro 1.

Quadro 1

Classificação das Técnicas de Previsão Segundo o Nível de Agregação dos Dados

| | Nível Individual | Nível Agregado |
|---------------------|--|---|
| Produtos existentes | Modelos de varredura de dados do comportamento individual (Tipo I) | Modelos econométricos de comportamento do mercado (Tipo II) |
| Novos produtos | Modelos de teste e pré-teste de mercado (Tipo III) | Modelos de difusão da inovação (Tipo IV) |

Fonte: Roberts (1998).

Dentro dessa classificação proposta por Roberts (1998), o trabalho aqui apresentado localiza-se na categoria de previsão modelagens Tipo I, pois considera a sensibilidade das vendas de um produto às variáveis do composto de *marketing*.

Esse tipo de análise pode ser utilizado para obter o composto de *marketing* otimizado (que resulta em maior faturamento ou lucro, por exemplo) em ambientes altamente competitivos, para entender fontes de heterogeneidade de consumidores e para estudar as relações entre o ponto-de-venda e a indústria em diferentes segmentos (ROBERTS, 1998).

Diversas técnicas têm sido utilizadas nesse tipo de modelagem, entre elas modelos de escolha discreta, árvores de decisão, técnicas de regressão linear múltipla e análises de séries temporais. Todas elas demandam um grande esforço do analista para que possam ser adaptadas para trabalharem com dados individualizados. Os resultados nem sempre são satisfatórios (ROBERTS, 1998).

Uma das técnicas para previsão desagregada que vem sendo utilizada com sucesso é a modelagem com uso de redes neurais artificiais, tendo mostrado superioridade sobre várias outras modelagens, em diversas áreas do conhecimento (ANSUJ *et al.*, 1996).

Modelos econométricos e de análise de séries de tempo enquadram-se nos modelos do Tipo II. Essa classe de modelagem foi revista por Hanssens, Parsons e Schultz (1990), e aplicações e exemplos das técnicas podem ser encontrados em Brown (1959), Klein e Young (1980) e Gross e Peterson (1983). Exemplos típicos da utilização desses modelos englobam a

previsão de tendências de mercado, como a taxa de retração de uma indústria ou setor; de índices macroeconômicos, como inflação e crescimento do Produto Interno Bruto (PIB); de vendas agregadas em uma empresa, como o total de demanda anual ou mensal; dos efeitos de variáveis do composto de *marketing* na demanda total: exame de canais de distribuição entre outros.

Modelos do Tipo III são calibrados com dados de pré-lançamento, utilizando, para tanto, dados individualizados de respondentes selecionados. Os modelos assim desenvolvidos são geralmente voltados ao estudo da participação e penetração no mercado para novos produtos. Outras possibilidades incluem a estimação do tamanho de um novo mercado ou de sua transformação com a entrada de um novo produto ou marca (ROBERTS, 1998).

Esses modelos são empregados na seleção de cesta de produtos para venda casada (BEN-AKIVA e GERSHENFELD, 1998).

Os modelos do Tipo IV, voltados para previsão de difusão de inovações, surgiram na década de 1960 com Fourt e Woodlock (1960) e Mansfield (1961). Bass (1969) propôs, então, um modelo de difusão de inovação que se tornou referência na literatura (MAHAJAN, BASS E MULLER, 1990). Revisões do seu modelo têm sido usadas para prever a difusão de inovações em varejo, tecnologia industrial, agricultura, educação, farmácia e mercados de bens duráveis. O objetivo de um modelo de difusão é prever a intensidade de adoção de uma inovação ao longo do tempo no mercado. Mahajan, Bass e Muller (1990) trazem uma revisão completa dos modelos de inovação e seus avanços.

2.2. Seleção de variáveis preditivas e o composto de marketing

O segundo aspecto fundamental do processo de previsão de vendas é, neste estudo, a escolha das variáveis preditivas que permitem estimar as vendas futuras de um produto.

Explorou-se a previsão de vendas a partir dos componentes do composto de *marketing* do varejo. Parente (2000, p.61) define composto de *marketing* como “o conjunto de variáveis que compõem o esforço de *marketing* do varejista e engloba todos os fatores controláveis que o varejista articula [...] para conquistar as preferências dos consumidores”. De acordo com o autor, a fatia de mercado de uma loja depende principalmente das características de seu composto de *marketing* ante o composto de *marketing* de seus concorrentes.

O composto de *marketing* do varejo pode ser classificado a partir de seis macrofatores (MASON, MAYER e WILKINSON, 1993) chamados **seis Ps**: *mix* de produtos, preço, promoção, apresentação, pessoal, ponto-de-venda.

Segundo Mason, Mayer e Wilkinson (1993), o comportamento de compra do consumidor em uma loja é influenciado por variáveis ambientais (economia, clima, demografia, época

do ano, tendências de mercado), pelo composto de *marketing* da loja e pelo composto de *marketing* da concorrência. Influenciado por esses três grupos de forças, o consumidor decide o quanto gastar.

Este trabalho concentra-se na exploração do impacto na demanda proporcionado por alterações nos três primeiros componentes do composto de *marketing*: produto, preço e promoção. Essa escolha deve-se à dificuldade de obtenção de dados representativos dos outros três componentes: apresentação, pessoal e ponto-de-venda.

2.3. Seleção de produtos afins — identificação de centro de interesse

Ainda que se procure buscar a relação entre produtos e a influência cruzada em suas vendas, o grande número de itens de produtos em venda em uma loja de varejo torna necessária a utilização de um critério de seleção de um número tratável de produtos a considerar em conjunto. Não seria viável tratar em um só modelo todos os produtos de uma loja (48.000 itens no caso desta pesquisa).

Para fazer a seleção de produtos para o modelo de previsão, pode-se utilizar a técnica de análise de cesto de compras (BERRY e LYNOFF, 1997). Essa análise permite selecionar um conjunto de produtos que possuem uma correlação em suas vendas. Pode-se, assim, trabalhar com esse conjunto de produtos que forma um **centro de interesse** (ou categoria de produtos) nos modelos de previsão. Busca-se, então, a previsão de demanda dentro desse centro de interesse em particular.

A partir da análise do cesto de compras, são geradas regras do tipo “se o consumidor compra o produto A, há uma probabilidade P de que ele também compre o produto B”. A análise é executada por transação e utiliza duas medidas essenciais: o suporte e a confiança. O suporte representa a frequência com que a regra aparece quando comparada ao total de transações efetuadas. Caso a regra apareça, por exemplo, em 200 do total de 500 transações, o suporte será de 40% (200/500). A confiança representa a frequência com que um cliente que compra o produto A, também compra o produto B. Se todos os clientes que compraram o produto A também compraram B, a confiança é de 100%.

Um centro de interesse é um grupo de produtos que estão de alguma forma relacionados. Esse relacionamento pode ser medido pela elasticidade cruzada entre os produtos, que reflete como a variação do preço de um produto afeta o comportamento de vendas de outros (PARENTE, 2000). Havendo relacionamento na forma de elasticidade cruzada entre um grupo de produtos, conclui-se que tais produtos formam um centro de interesse e que dois produtos quaisquer desse grupo devem ser, necessariamente, complementares ou substitutos. É, portanto, a relação entre a variação da demanda do produto A e a variação do preço do produto B (correlação preço-demanda):

$$\text{Elasticidade Cruzada} = \frac{\Delta Q_a}{\Delta P_b}$$

Em outras palavras, se há elasticidade cruzada, deve existir sempre uma correlação entre o preço e a demanda de dois produtos pertencentes à mesma categoria. Se essa correlação for negativa, os produtos são complementares. Sendo a correlação positiva, os produtos são substitutos.

3. REDES NEURAIS ARTIFICIAIS

3.1. Redes neurais aplicadas à previsão

Segundo Berry e Linoff (1997), a modelagem por meio de redes neurais artificiais é uma das técnicas de mineração de dados mais utilizada, sendo largamente empregada em tarefas de classificação e previsão. O desempenho dos computadores permite, atualmente, que se processem grandes bases de dados por meio de redes neurais (DETIENNE, DETIENNE e JOSHI, 2003). Neste trabalho foram utilizadas redes neurais para a previsão de vendas.

Métodos estatísticos de séries temporais, tais como médias móveis ou modelos auto-regressivos, têm sido apresentados na literatura para previsão de vendas. No entanto, uma vez que baseiam a previsão de vendas exclusivamente em seus valores históricos, esses modelos falham quando o desempenho das vendas se baseia em outras variáveis exógenas como variações de preços, campanhas e influência da mídia. Técnicas mais flexíveis, como redes neurais, permitem melhor desempenho de previsão (FRANK *et al.*, 2003).

O desempenho das redes neurais é destacado em diversos estudos, como é apresentado a seguir.

Segundo Detienne, Detienne e Joshi (2003), um atrativo das redes neurais surge a partir das limitações de técnicas convencionais de regressão múltipla, tais como incapacidade de lidar com não-linearidade, exigência de estabelecimento de um modelo subjacente preestabelecido, sensibilidade a multicolinearidade, exigências quanto à distribuição dos dados, entre outras limitações.

Diversos autores mostraram a capacidade das redes neurais em lidar com fenômenos que têm comportamentos não-lineares (LUXHOJ, RIIS e STENSALLE, 1996; ZHANG, PATUWO e HU, 1998). Essa é uma característica primordial, já que os problemas em *marketing* sempre lidam com dados como vendas e preços, em que as séries de dados apresentam pontos de inflexão, tendências e não-linearidades (BAETS e VENGOPAL, 1994). Algumas vezes os dados podem ter comportamento caótico (THALL, 1992; HIBBERT e WILKINSON, 1994) e, mesmo nessas condições, podem ser feitas previsões com precisão por uma rede neural (LAPEDES e FARBER, 1987).

As redes neurais não exigem que se conheça com antecedência a forma de distribuição dos dados (BAETS e VENGU-

OPAL, 1994; ZHANG, PATUWO e HU, 1998; WRAY, BEJOU e PALMER, 1994). Podem, automaticamente, ignorar dados que não contribuem para a qualidade preditiva do modelo (CHURCH e CURRAM, 1996). Segundo Baets e Venugopal (1994), as redes neurais são robustas e tolerantes a inconsistências nos dados.

Pode-se demonstrar que uma rede neural é capaz de aproximar qualquer função contínua (ZHANG, PATUWO e HU, 1998). Com uma rede neural pode-se também modelar uma série temporal qualquer (ANSUJ *et al.*, 1996).

Relações de multicolinearidade são automaticamente identificadas e assimiladas pela rede, conforme Wray, Bejou e Palmer (1994). Segundo os autores, o desempenho das redes neurais não é depreciado em virtude da ocorrência de multicolinearidade, da ausência de normalidade ou da existência de *outliers* nos dados, ao contrário das técnicas tradicionais de previsão, como regressão ou análise discriminante, que impõem para seu uso hipóteses de normalidade nos dados e ausência de multicolinearidade.

3.2. O método de aprendizado por retropropagação

Na pesquisa aqui relatada foram criadas redes segundo o modelo de multicamadas e o método de aprendizado de retropropagação proposto por Rumelhart, Hinton e Williams (1986), mais comumente utilizado para treinamento de redes neurais (DETIENNE, DETIENNE e JOSHI, 2003) e problemas de classificação e identificação de padrões nos dados (META-XIOTIS e PSARRAS, 2004). Nesse tipo de rede, os neurônios artificiais são organizados em camadas horizontalmente (figura 1). Cada neurônio conecta-se e envia informação para todos os neurônios da camada seguinte. Neurônios pertencentes à mesma camada não são interligados. Essas redes são frequentemente constituídas de três camadas, uma vez que qualquer função contínua pode ser aproximada por uma rede com três camadas (ZHANG, PATUWO e HU, 1998): a camada de entrada, com os neurônios de entrada; a camada intermediária, com os neurônios intermediários; e a camada de saída, com o neurônio de saída. Os neurônios são unidades básicas que processam os dados introduzidos na rede neural. Os dados são processados e propagados na rede por meio de equações matemáticas, chamadas equações de aprendizado.

Pelos neurônios de entrada introduz-se informação na rede. O neurônio de saída transmite as respostas da rede. Os neurônios transmitem a informação a partir dos dados de entrada, por meio de conexões determinadas a partir de um processo chamado de processo de aprendizado. São determinados os pesos de conexão entre os neurônios. Os pesos de conexão permitem aos neurônios ponderar a influência das variáveis preditivas e dos dados de entrada na rede e na resposta que ela gera. Uma variável preditiva com forte influência no resultado a ser obtido fará com que os pesos de conexão entre a variável e os neurônios a ela conectados sejam maiores. Se, por exem-

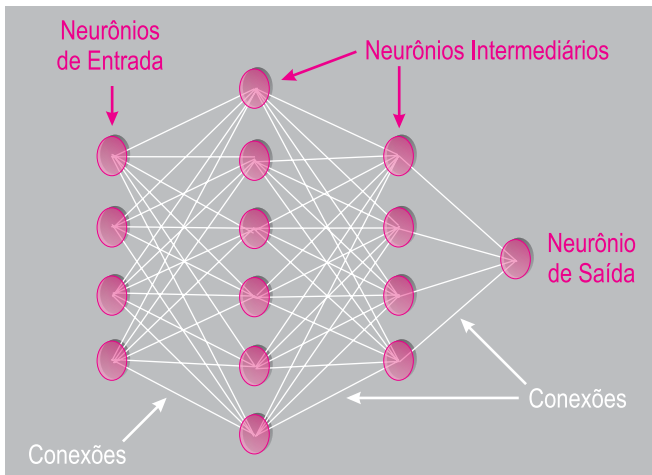


Figura 1: Modelo de Rede Neural Multicamada

plo, o preço do produto tem forte influência na previsão de vendas, as conexões dos neurônios ligados ao preço terão maior valor. Se, ao contrário, o preço pouco influencia a venda, os pesos de conexão dos neurônios ligados ao preço terão pequeno valor.

Os pesos são determinados no processo de aprendizado de maneira que, a partir dos dados de entrada, as respostas desejadas sejam obtidas no neurônio de saída. Neurônios e camadas intermediárias são colocados entre as camadas de entrada e saída a fim de permitir melhor capacidade de separação dos dados e, em consequência, maior capacidade de previsão.

Neste estudo, usa-se um modelo de retropropagação baseado nos Modelos de Processamento Distribuído Paralelo propostos por Rumelhart e McClelland (1986).

A propagação de informação pela rede é efetuada como segue (figura 2): os valores de entrada são transmitidos de uma camada para a outra e transformados por meio de pesos de conexões entre os neurônios (W_{ji} : peso da conexão entre o neurônio I e o neurônio J).

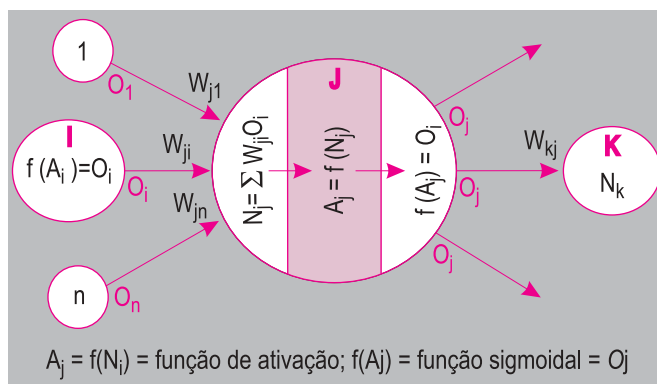


Figura 2: Propagação de Sinais na Rede Neural entre Neurônios das Camadas I, J, K

Fonte: Almeida e Dumontier (1996).

Duas equações matemáticas principais regem o processo de transmissão da informação no neurônio (equação [1]) e ajuste dos pesos de conexão (equação [2]) visando à ponderação adequada dos dados propagados ao longo da rede.

Pode-se tomar como exemplo um neurônio J na camada j. A intensidade do sinal enviado de um neurônio para o outro é uma função da intensidade do valor de ativação, gerado por uma função de ativação $f(N_j)$. Uma função de transferência leva em conta o valor de ativação e produz o sinal de saída O_j .

$$O_j(t+1) = \frac{1}{1 + e^{-(\sum W_i O_i + \theta_j)}} \quad [1]$$

A correção dos pesos de conexão no processo de aprendizado é feita de maneira iterativa e é dada pela equação [2]. Os dados (por exemplo, dados de vendas de produtos) são introduzidos repetidas vezes na rede. Cada dado introduzido é propagado na rede pela equação [1]. O erro entre o valor de saída desejado e o valor gerado de cada neurônio é computado (δ_j). Um novo peso de conexão $W_{ji}(n+1)$ é calculado a partir da equação [2]. O parâmetro μ determina com que intensidade o erro irá influenciar no cálculo do novo valor do peso, e o parâmetro α determina com que intensidade o valor anterior do peso irá influenciar no novo valor do peso da conexão ji .

$$\Delta W_{ji}(n+1) = \mu \delta_j O_i + \alpha \Delta W_{ji}(n) \quad [2]$$

3.3. Avaliação do desempenho das redes neurais

A fim de avaliar seu desempenho, neste estudo as redes neurais foram confrontadas com as modelagens *naïve* e de regressão linear. Tais técnicas foram escolhidas por serem bastante utilizadas para a determinação da capacidade preditiva mínima de modelos de previsão (MAKRIDAKIS, WHEELWRIGHT e MCGEE, 1983) e amplamente usadas na prática pelas empresas (HUSSEY e HOOLEY, 1995).

Um modelo de previsão *naïve* não-ajustado é aquele em que se toma o último registro conhecido como sendo a previsão para o próximo período, sem considerar a sazonalidade da variável dependente. Toma-se, por exemplo, o valor de venda na semana anterior para prever as vendas na semana seguinte. A comparação com a previsão *naïve* é uma referência comum encontrada na literatura para medir o ganho de desempenho ao se adotarem modelagens mais sofisticadas (LAWRENCE, EDMUNDSON e O'CONNOR, 2000). É também o método de previsão atualmente utilizado na empresa em que os dados foram coletados.

A fim de avaliar o desempenho (acurácia) absoluto e relativo dos diversos modelos de previsão, diversas ferramentas matemáticas foram desenvolvidas com o tempo. Nesta pesquisa foram utilizadas duas medidas comumente empregadas na avaliação do desempenho das técnicas de previsão (ACZEL, 1993):

- *Mean Square Error* (MSE) ou Erro Quadrado Médio — é uma medida que calcula a média dos erros observados entre o valor gerado pela rede neural, isto é, a previsão da rede e o valor desejado (o valor observado na prática).

$$MSE = \frac{\sum_{i=1}^n ei^2}{n} \quad [3]$$

- *Mean Absolute Percentage Error* (MAPE) ou Erro Percentual Absoluto Médio — é uma medida do erro absoluto médio em termos percentuais, para que se tenha uma visão do erro comparado com o valor previsto. É também usada por permitir comparações entre modelos que utilizam dados diferentes.

$$MAPE = 100 \cdot \frac{\sum_{i=1}^n \left| \frac{Xi - Fi}{Xi} \right|}{n} \quad [4]$$

4. MÉTODO DE PESQUISA

A presente pesquisa é do tipo experimental, em que se busca avaliar a capacidade preditiva das técnicas empregadas.

Construiu-se, por meio do uso de redes neurais artificiais, um modelo de previsão de vendas a fim de prever a demanda individualizada de produtos, levando em consideração fatores explicativos da demanda (modelagem causal).

A exploração dos dados, a construção e a avaliação dos modelos criados seguiram os seguintes passos (figura 3):

- obtenção dos dados;
- seleção e limpeza dos dados;
- análise de cesto de compras para seleção de um grupo de produtos relacionados entre si (centro de interesse do cliente);
- tratamento dos dados — preparação dos dados do centro de interesse selecionado para a criação dos modelos de redes neurais;
- criação e teste dos modelos de previsão de vendas baseados em redes neurais;
- comparação com outras técnicas — *naïve* e regressão linear.

A fim de proceder ao tratamento dos dados, à identificação dos centros de interesse e à criação dos modelos de previsão, foram utilizados *software* de *data mining*. Além do SPSS, *software* mais tradicional para o tratamento inicial dos dados, foram utilizados o pacote *Intelligent Miner* da IBM, o *Mega-puter Poly Analyst 4.5* e o *Statistica Neural Networks* para a criação das redes neurais.

5. DESENVOLVIMENTO E RESULTADOS DA PESQUISA

A pesquisa foi desenvolvida com base nas vendas de uma empresa varejista brasileira. Como se trata de uma rede varejista de grande porte, os produtos comercializados por ela são bastante variados, incluindo eletroeletrônicos, utilidades domésticas, móveis, vestuário, brinquedos, entre outros. A rede de pontos-de-venda é distribuída em todo o território nacional, totalizando perto de 100.000 metros quadrados de área de venda distribuídos em mais de 100 pontos-de-venda.

Diferentes configurações de dados, seleções de variáveis e tratamentos nos dados foram efetuados a fim de identificar a configuração com melhor desempenho. As diversas redes testadas foram avaliadas com base nas medidas de erro MSE e MAPE, conforme exposto anteriormente, permitindo avaliar o desempenho das diferentes configurações. Os resultados do teste com 144 redes diferentes indicaram melhor desempenho das redes em que foram feitas normalizações (BERRY e LYNOFF, 1997; HAND, 1998) e padronizações nos dados (WRAY, BEJOU e PALMER, 1994), bem como supressão de variáveis correlacionadas (HAIR JR. *et al.*, 1998). Melhor desempenho também foi observado quando se utilizaram as vendas passadas de produtos inter-relacionados em um mesmo centro de interesse do consumidor, isto é, vendas de grupos de produtos

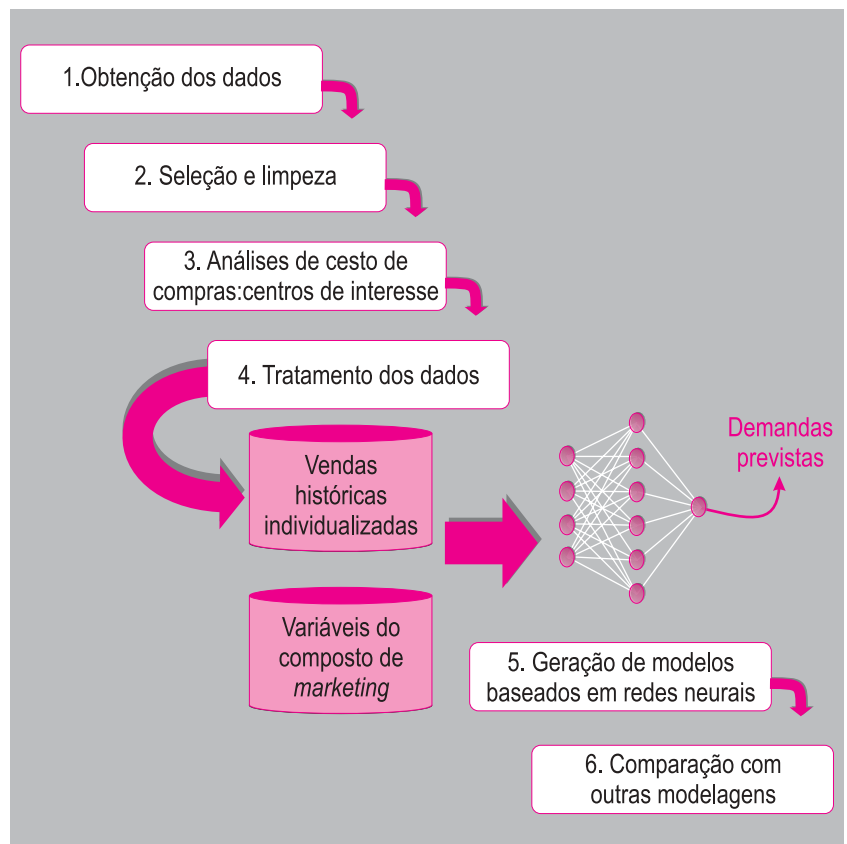


Figura 3: Etapas para a Exploração dos Modelos de Previsão

com vendas inter-relacionadas. No item 5.1., é exposta a forma de seleção da amostra. O item 5.2., análise de dados, é uma etapa de análise exploratória dos dados, na qual se buscou avaliar o impacto de transformações nos dados, de correlação entre variáveis e da inclusão ou exclusão de variáveis preditivas inicialmente selecionadas no desempenho das redes neurais. No item 5.3., são mostrados os resultados de previsão das redes em períodos de tempo à frente.

5.1. Seleção da amostra e variáveis

5.1.1. Seleção de dados de uma loja

Foram coletados dados relativos às vendas da empresa no período de novembro de 1998 a junho de 2002, totalizando 43 meses. Os dados são referentes a todos os produtos vendidos em cinco pontos-de-venda selecionados pela empresa segundo seu interesse. Esses pontos foram selecionados por representarem tipos distintos de lojas (tamanho, localização, *mix* de produtos etc.). Em razão do volume de dados e das limitações de tratamento desses dados, das cinco lojas apenas uma foi retida para a realização do estudo. Foi escolhida a loja com maior volume de transações com o objetivo de obter-se maior diversidade de itens comercializados.

Foram selecionadas apenas as vendas realizadas a prazo (60% das vendas) da loja selecionada. São vendas identificadas, em que há informações precisas sobre o cliente, permitindo análises do comportamento de consumo por cliente.

Inconsistências nos registros, como datas de vendas inconsistentes, juros negativos, estoques sem registro, ausência de classificação de produto, entre outras, obrigaram à eliminação de certo número de registros. No quadro 2 indica-se o tamanho final da amostra.

Quadro 2

Amostra Utilizada na Pesquisa

| | |
|------------------|---------|
| Registros | 12.560 |
| Produtos | 881 |
| Marcas Distintas | 558 |
| Periodicidade | Semanal |
| Vendas | A prazo |

5.1.2. Análise do cesto de compras

A utilização da técnica de análise do cesto de compras teve como propósitos:

- separar um grupo coeso de produtos, minimizando o total de produtos a serem previstos pelo modelo, uma vez que o objetivo da pesquisa foi testar a aplicabilidade das redes neurais e da metodologia de modelagem empregada, e não construir um modelo de previsão para todos os produtos da empresa estudada;

- determinar quais produtos fazem parte de um mesmo centro de interesse e, assim, investigar se essa informação traz alguma melhoria na capacidade preditiva de modelos de previsão baseados em redes.

Dois estratégias de abordagem dos dados foram tomadas: análise por meio das transações, buscando quais itens são vendidos em conjunto numa mesma transação; e análise horizontal por cliente, procurando quais produtos um mesmo cliente compra ao longo do tempo. A partir do *software* utilizado, *Intelligent Miner 6.1* (IBM, 1999), o centro de interesse foi identificado visualmente pelas relações apresentadas graficamente pelo *software*. Os relacionamentos foram plotados pelo *software*, a partir das medidas de relacionamento apresentadas anteriormente. Um processo iterativo permite ao pesquisador identificar e destacar as relações mais significativas.

Uma vez que o objetivo neste artigo não é explorar a identificação de cesto de compras, mas sim a capacidade preditiva das redes neurais, optou-se por não detalhar aqui esses resultados, o que poderá ser objeto de publicações futuras. A inspeção visual (ver nota no final do corpo do artigo) mostrou que as linhas de produtos apresentadas no quadro 3 estão conectadas, formando um grupo coeso e, conseqüentemente, um centro de interesse. Tendo sido identificado esse centro de interesse pela análise de cesto de compras, seus dados foram utilizados na exploração e uso de redes neurais.

Quadro 3

Centro de Interesse Detectado

| Linhas de Produtos | Código |
|---------------------|-----------------------------------|
| Ventiladores | 204 |
| Processadores | 212 |
| Liquidificadores | 217 |
| Ferros de passar | 219 |
| Batedeiras | 237 |
| Telefonia | 1.300 |
| Panelas | 1.443 |
| Brinquedos diversos | 1.805, 1.810, 1.817, 1.826, 1.838 |

5.1.3. Seleção de variáveis

As variáveis independentes utilizadas para a previsão de vendas são apresentadas no quadro 4. Frank *et al.* (2003) sugerem o interesse em utilizar variáveis exógenas na previsão de vendas, tais como preço, propaganda, divulgação na mídia, além das variáveis tradicionalmente utilizadas em previsão de séries temporais, como média móvel e valores históricos de vendas dos produtos (BARASH e MITCHEL, 1998; GORDON, 1998; DeLURIGO, 1998; ALIBAIG e LILLY, 1999).

A inclusão das variáveis baseou-se em sua disponibilidade na base de dados. Não havendo um critério conceitual para

sua escolha, procurou-se incluir todas as variáveis que eventualmente tivessem relação com as variáveis sugeridas na literatura. Sua influência foi então testada a partir da criação e do teste das diversas configurações de redes neurais.

Quadro 4

Variáveis Independentes Usadas nos Modelos de Previsão Semanal

| Variável | Utilização |
|--|---|
| Semana | Variável de tempo, contabilizando as semanas seqüencialmente. A contagem inicia-se em 01 de novembro de 1998, considerada como semana 1. A última semana, que se inicia no dia 26 de maio de 2002, é a semana 187. |
| Preço Médio Semanal | Média de preço praticado por produto, durante a semana correspondente. |
| Juros Médios Semanais | Taxa de juros média praticada por produto, durante a semana correspondente. |
| Prazo | Porcentagem de vendas realizadas a prazo na semana anterior. |
| Id_produto, Produto, Linha, Setor, Id_marca | Variáveis de identificação do produto. |
| Classe, Classe_distribuição | A empresa adota duas classificações distintas para um mesmo produto: uma baseada no giro do produto, e outra baseada no tipo de compra/distribuição a que ele está submetido. Ambas, entretanto, identificam parâmetros de alta, média ou baixa venda, venda sob encomenda, venda sem mostruário, venda direta do fornecedor e produto fora de linha. |
| Comissão | Indica se o produto entra na comissão dos vendedores na semana correspondente. |
| Estoque, Estoque_dias, Reposição_dias, Mostruário, Grade | Representam a disponibilidade do produto, seja em estoque, seja em mostruário, seja em prazo de entrega/reposição. |
| Campanha, Campanha_agressiva, Veiculações | Variáveis que medem o esforço promocional efetuado na semana. Para esforços corporativos, todos os produtos recebem o mesmo valor para a variável Campanha no período correspondente. As campanhas agressivas referem-se a apenas alguns grupos de produtos, enquanto Veiculações mede a exposição das campanhas na mídia. |
| Variáveis com lag de tempo | Foram testadas variáveis com lag de tempo para as condições em que se detectaram correlações horizontais. |

Utilizaram-se, num primeiro momento, os dados em suas escalas originais, já que a primeira análise realizada foi a do cesto de compras, que buscou apenas o estudo das frequências dos registros de itens vendidos para detectar os produtos inter-relacionados, sinalizando os centros de interesse do consumidor.

Os dados originais foram transformados a fim de aproximar sua distribuição da normal e melhorar a assimilação dos dados por modelos quantitativos (BERRY e LYNOFF, 1997; HAND, 1998).

5.2. Análise de dados

5.2.1. Transformação e padronização dos dados

Todas as variáveis, com exceção de “Juros Médios Semanais”, apresentaram comportamentos não-desejados, com distribuições muito distantes da normal. As transformações necessárias foram obtidas por meio do gráfico dispersão *versus* nível ($\text{Log}M \times \text{Log}D_f$), de acordo com o procedimento descrito por Tukey, Hoaglin e Mosteller (1983), conforme a equação [5].

$$\text{Log } D_f = k + b.\text{Log } M \quad [5]$$

em que:

M é a mediana da variável tomada — nesse caso, por linha de produto;

D_f é a medida do interquartil da mesma variável.

Uma vez traçada a melhor reta para a equação [5], é aproximado o valor de b , que indica a melhor transformação nos dados, segundo a equação [6].

$$\text{Transformação: } Z = x^{1-b} \quad [6]$$

A tabela 1 mostra os resultados obtidos sem transformação e com transformação dos dados. Pode-se notar que o MSE dos dados transformados (MSE = 1,93) é bastante inferior ao MSE dos dados não transformados (MSE = 17,5). Confirmou-se, assim, o interesse em transformar os dados.

Foi também efetuada a padronização dos dados (tabela 2), colocando-os todos em uma mesma escala (média 0 e desvio-padrão 1), visando à melhoria da capacidade preditiva das redes (WRAY, BEJOU e PALMER, 1994). Pode-se notar que o erro MSE diminuiu ao se compararem os erros sem padronização (MSE = 1,9306) e com padronização (MSE = 1,9108).

5.2.2. Eliminação de variáveis correlacionadas

O desempenho de previsão de técnicas estatísticas convencionais é fortemente influenciado pela correlação existente entre as variáveis explicativas. A inclusão de variáveis correlacionadas em um modelo torna os coeficientes da equação

Tabela 1

Influência da Transformação dos Dados na Capacidade Preditiva das Redes Neurais

| Rede | Variáveis de Entrada | Variável Prevista | Nós | Camadas | R ² | MSE |
|---------------------|---|-------------------|-----|---------|----------------|---------|
| 1- Não transformada | Id_produto, Comissão, Setor, Linha, Id_marca, Classe, Classe_distribuição, Estoque, Estoque_dias, Reposição_dias, Mostruário, Grade, Semana, PrecoMédioSemanal, JurosMédiosSemanais, Pprazo | Venda7 | 3 | 3 | 0,2194 | 17,5000 |
| 2 - Transformada | Id_produto, Comissão, Setor, Linha, Id_marca, Classe, Classe_distribuição, Estoque_log, Estoque_dias-2, Reposição_dias, Mostruário_log, Grade_log, Semana, Preco_pot, JurosMédiosSemanais, Pprazo | Vendas_log | 14 | 3 | 0,6052 | 1,9306 |

Tabela 2

Resultados da Rede Neural 10, com as Entradas Numéricas Normalizadas e Padronizadas

| Rede | Variáveis de Entrada | Variável Prevista | Nós | Camadas | R ² | MSE |
|------|---|-------------------|-----|---------|----------------|--------|
| 10 | Id_produto, Comissão, Setor, Linha, Id_marca, Classe, Classe_distribuição, Estoque_logP, Estoque_dias-2P, Reposição_diasP, Semana, Preco_potP, JurosMédiosSemanaisP, PprazoP, Mostruário_logP | Vendas_log | 26 | 4 | 0,6176 | 1,9108 |

instáveis (HAIR JR. *et al.*, 1998). Nesta pesquisa, foi testada a influência da inclusão de variáveis fortemente correlacionadas (tabela 3) no modelo de redes neurais.

Tabela 3

Correlações Significativas Detectadas entre as Variáveis Independentes

| Correlações entre Variáveis Independentes | Correlação de Pearson | Nível de Significância |
|---|-----------------------|------------------------|
| Estoque_log ↔ Grade_log | 0,590 | 0,01 |
| Mostruário_log ↔ Preco_pot | -0,459 | 0,01 |
| Classe_distribuição ↔ Classe | 0,556 | 0,01 |

Pôde-se constatar redução no erro MSE com a eliminação de variáveis correlacionadas. A melhora mais significativa foi

obtida com a eliminação da variável Grade_log (tabela 4).

5.2.3. Exploração de variáveis de entrada

O impacto de diferentes variáveis independentes foi avaliado na capacidade preditiva das redes. Foram geradas 66 redes neurais com a incorporação das variáveis apresentadas no quadro 5. Todas utilizaram o mesmo conjunto de dados, sempre prevendo para uma semana à frente e utilizando entradas numéricas normalizadas e padronizadas.

Em resumo, de todas as variáveis incorporadas ao modelo visando ampliar as informações disponíveis sobre o composto de *marketing* e sobre a sazonalidade, apenas a informação sobre o total do produto vendido na semana anterior interferiu significativamente na capacidade preditiva da rede neural, conforme MSE de 1,8441 da rede 26 (tabela 5).

Tabela 4

Resultados da Rede Neural 6, Sem a Presença da Variável "Grade"

| Rede | Variáveis de Entrada | Variável Prevista | Nós | Camadas | R ² | MSE |
|------|--|-------------------|-----|---------|----------------|--------|
| 6 | Id_produto, Comissão, Setor, Linha, Id_marca, Classe, Classe_distribuição, Estoque_log, Estoque_dias-2, Reposicao_dias, Mostruário_log, Semana, Preco_pot, JurosMédiosSemanais, Pprazo | Vendas_log | 3 | 3 | 0,6127 | 1,9187 |

Quadro 5
Exploração das Variáveis de Entrada

| Variáveis Exploradas | Resultados Obtidos |
|--|--|
| <p>Dados sobre a sazonalidade e a passagem do tempo</p> <ul style="list-style-type: none"> Semana_ano, representando a semana do ano, variando de 1 a 52 Mês_ano, representando o mês do ano, variando de 1 a 12 12 entradas binárias (0 ou 1), representando os meses do ano 52 entradas binárias (0 ou 1), representando as semanas do ano | Não houve melhorias. A rede com melhor resultado foi aquela com a medida original e isolada da variável Semana. |
| <p>Análise de lags de tempo das vendas</p> <ul style="list-style-type: none"> Vendas da semana anterior Vendas da 2ª semana anterior Vendas da 10ª semana anterior Vendas da 42ª semana anterior Vendas da 52ª semana anterior Vendas das 52ª, 42ª, 10ª, 2ª semanas e semana imediatamente anterior Todos os casos anteriores conjugados Todos os casos anteriores, preenchendo-se valores <i>missing</i> | Os resultados mostraram que houve melhoria com a adoção das variáveis de vendas com <i>lag</i> de tempo, desde que os <i>missing values</i> fossem preenchidos. Nesse caso, o valor adotado nos registros faltantes foi igual ao último (próximo no tempo). Ressalta-se que a melhoria de desempenho foi visível apenas nos <i>lags</i> de 1 e 2 semanas, sendo muito mais pronunciado com o uso apenas do <i>lag</i> de 1 semana (Rede 26 — Tabela 5), e sendo indiferente a adoção de ambos. |
| <p>Análise de lags de tempo das variáveis independentes</p> <ul style="list-style-type: none"> Grade do produto (quantidade especificada a ser mantida por loja) da semana anterior Estoque na semana anterior Preço na 2ª semana anterior Preço na 3ª semana anterior Juros médios na 4ª semana anterior Todos os casos anteriores, preenchendo-se valores <i>missing</i> | Não houve resultados positivos nos modelos gerados com o uso dos <i>lags</i> cruzados. |
| <p>Incorporação de dados de promoção (esforço de <i>marketing</i>)</p> <ul style="list-style-type: none"> Dados de campanhas promocionais institucionais, atribuídos com valores 0 ou 1 para todos os produtos Dados de campanhas de <i>marketing</i> agressivas, direcionadas somente a alguns produtos Número de veiculações na mídia Todos os casos anteriores conjugados | As variáveis promocionais não trouxeram melhora na capacidade preditiva das redes. A explicação mais provável é que houve dificuldade em atribuir a um produto a campanha realizada, já que os registros de campanhas realizadas não eram completos nem precisos. Isso obrigou a atribuir a campanha a toda uma linha de produtos, ou mesmo setor, na maior parte dos casos. |

Tabela 5
Rede com a Inclusão do Total de Produtos Vendidos na Semana Anterior

| Rede | Variáveis de Entrada | Variável Prevista | Nós | Camadas | R ² | MSE |
|------|--|-------------------|-----|---------|----------------|--------|
| 26 | Id_produto, Comissão, Setor, Linha, Id_marca, Classe, Classe_distribuição, Estoque_logP, Estoque_dias-2P, Reposição_diasP, Semana, Preço_potP, JurosMédiosSemanaisP, PprazoP, Mostuário_logP, Lag1 | Vendas_log | 7 | 3 | 0,6585 | 1,8441 |

5.2.4. Incorporação de informações do centro de interesse

A fim de identificar se o modelo criado é capaz de incorporar informações sobre o relacionamento entre produtos que

fazem parte de um mesmo centro de interesse do consumidor — grupo de produtos inter-relacionados —, novas variáveis foram incorporadas, conforme consta no quadro 6. Os modelos criados e apresentados na sessão anterior não levaram em conta

Quadro 6

Incorporação de Informações de Centro de Interesse

| Variável | Rede Testada |
|---|-----------------|
| Vendas totais da linha de produtos correspondente | Rede 79 |
| Vendas totais do setor da loja correspondente | Rede 80 |
| Vendas totais de todo o centro de interesse | Rede 81 |
| Preço médio da linha de produtos correspondente | Rede 82 |
| Preço médio do setor da loja correspondente | Rede 83 |
| Preço médio de todo o centro de interesse | Rede 84 |
| Diferença de preço entre o produto e a média da linha de produto | Rede 85 |
| Diferença de preço entre o produto e a média do setor da loja | Rede 86 |
| Diferença de preço entre o produto e a média de todo o centro de interesse | Rede 87 |
| Idem, para entradas padronizadas | Redes 88 a 105 |
| Idem, com preenchimento de <i>missing values</i> assumindo o último valor existente | Redes 106 a 141 |

possíveis influências entre as vendas de diferentes produtos. As variáveis apresentadas no quadro 6 agregam dados de vendas de diferentes produtos com o intuito de representar o conhecimento sobre as elasticidades cruzadas entre os produtos.

Foi possível constatar grande melhora das redes ao se utilizar a informação de total de vendas dos produtos substitutos (pertencentes a uma mesma linha de produtos e, portanto, produtos concorrentes diretos). Não houve melhoria quando da utilização dos preços médios nem das diferenças de preços. A rede neural 108, cujos resultados são mostrados na tabela 6, demonstra que a incorporação de informações a respeito de produtos inter-relacionados pode trazer benefícios na previsão de vendas. O MSE caiu para 1,7457.

5.3. Exploração dos modelos de previsão

5.3.1. Previsão com redes neurais

A fim de avaliar a capacidade preditiva do modelo baseado em redes neurais, o conjunto de dados foi dividido em dois grupos:

- grupo de criação do modelo, constituído internamente por registros de treinamento e validação. Para a criação do modelo, foram utilizados os registros de novembro de 1998 a maio de 2001, num total de 135 semanas consecutivas, sempre prevendo para uma semana à frente;
- grupo de avaliação externo, simulando a utilização real do modelo de previsão em dados **desconhecidos**, nunca apresentados à rede neural. Os registros subsequentes à semana 135 (maio de 2001 a junho de 2002), totalizando 3.140 casos, foram usados para avaliar a capacidade preditiva do modelo gerado.

Foram criadas 141 redes com diferentes configurações de variáveis de entrada. A tabela 7 resume o processo de geração de redes neurais para previsão executado, indicando os critérios utilizados para a criação das diferentes redes. A rede com o

menor grau de erro na fase de teste anterior, rede 108, foi destacada das outras redes.

A rede de melhor desempenho, conforme consta na tabela 7, foi a 108, com o MSE de 1,7457. Essa rede incorpora informações sobre o centro de interesse, conforme apresentado a seguir.

Uma vez criadas as redes, o segundo subconjunto de dados (maio de 2001 a junho de 2002) foi utilizado para testá-las, dados esses independentes dos dados utilizados para a criação do modelo. A capacidade preditiva da melhor rede na fase de aprendizado (rede 108) foi testada com esse novo conjunto de dados.

O modelo baseado em redes neurais erra em sua previsão de vendas, em média, um total de 10,4 unidades de produto

Tabela 6

Resultados da Rede Neural 108, com Incorporação do Total de Vendas de Produtos Substitutos (Missing Values Preenchidos)

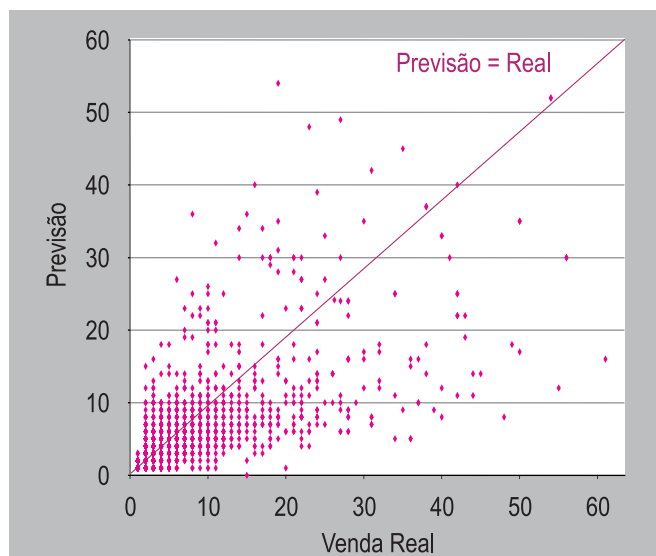
| Rede | Variáveis de Entrada | Variável Prevista | Nós | Camadas | R ² | MSE |
|------|---|-------------------|-----|---------|----------------|--------|
| 108 | Id_produto, Comissão, Setor, Linha, Id_marca, Classe, Classe_distribuição, Estoque_logP, Estoque_dias-2P, Reposição_diasP, Semana, Preço_potP, JurosMédiosSemanaisP, PprazoP, Mostuário_logP, Lag1, Vendas_linhaP | Vendas_log | 7 | 3 | 0,7164 | 1,7457 |

Tabela 7

Resumo das Redes Neurais Geradas e Desempenho Obtido

| Redes | Testes Realizados | Melhor Rede | MSE da Melhor Rede |
|----------|--|---|--------------------|
| 1 ~ 4 | Comparação entre entradas transformadas e não transformadas | Rede 2 — Todas as entradas transformadas | 1,9306 |
| 5 ~ 9 | Teste de retirada de variáveis correlacionadas | Rede 6 — Retirada isolada da variável "Grade" | 1,9187 |
| 10 ~ 11 | Comparação entre entradas padronizadas e não padronizadas | Rede 10 — Entradas numéricas padronizadas entre 0 e 1 | 1,9108 |
| 12 ~ 78 | Incorporação de novas variáveis para representar: sazonalidade, autocorrelações das vendas, correlações com <i>lags</i> de tempo, dados promocionais | Rede 26 — Incorporação das vendas da semana anterior (autocorrelação com <i>lag</i> de 1 semana) | 1,8441 |
| 79 ~ 141 | Incorporação de informações agregadas sobre o centro de interesse | Rede 108 — Incorporação do total de vendas de produtos substitutos (<i>missing values</i> preenchidos) | 1,7457 |

para previsão de uma semana à frente, o que corresponde ao acerto médio percentual de 41,21%. O gráfico a seguir mostra o desempenho da rede, comparando o valor previsto com o valor real. A fim de avaliar esse resultado, utilizaram-se a análise de regressão e o método *naïve* de previsão.



Vendas Previstas versus Vendas Reais

5.3.2. Análise de regressão

Para a análise de regressão, foram utilizadas as mesmas variáveis do Modelo 108, o qual obteve a melhor capacidade preditiva dentre aqueles testados. O procedimento de tratamento e separação também seguiu a utilização de dados normalizados e padronizados, e a separação em dois grupos de dados: dados de criação até a semana 135 e avaliação do desempenho com dados da semana 136 à semana 187. A equação de regres-

são obtida é composta pelas seguintes variáveis, utilizando o método *step wise* de seleção de variáveis (estatística *Wald*): Vendas da linha de produtos, Porcentagem de vendas a prazo, Vendas no período anterior, Estoque, Prazo de reposição.

Levando em consideração o grupo de dados de validação, a regressão gerou um MSE de 49,1%, superior ao erro de 41,21% das redes neurais.

Avaliando-se a equação de regressão, observa-se que ela possui erro de ordem 32% maior do que o do modelo baseado em redes neurais, comparando-se o erro quadrado médio, e 19,1% maior, em termos percentuais.

5.3.3. Método *naïve*

Um modelo de previsão *naïve* não-ajustado é aquele em que se toma o último registro conhecido como sendo a previsão para o próximo período, sem considerar a sazonalidade da variável dependente. A comparação com a previsão *naïve* é uma referência comum encontrada na literatura referente à medida do ganho de desempenho ao se adotarem modelagens mais sofisticadas (LAWRENCE, EDMUNDSON e O'CONNOR, 2000). Ela também é importante, já que é o método de previsão atualmente utilizado na empresa em que os dados foram tomados.

Observou-se que a modelagem baseada em redes neurais artificiais obteve desempenho 54% superior ao da previsão *naïve*, tomando-se por base a raiz do erro quadrado médio (RMSE), e 64% superior, se for considerado o erro médio percentual (MAPE), ponderando os resultados para os dados de validação.

5.3.4. Resumo dos resultados

Na tabela 8, apresenta-se de maneira comparativa a capacidade de previsão de cada uma das técnicas. Pode-se perceber

Tabela 8
Capacidade Preditiva dos Modelos de Previsão

| Modelo | Dados | MSE | RMSE | MAPE% |
|--------------|---|--------|-------|--------|
| Rede 108 | Dados de criação (até semana 135) | 55,84 | 10,33 | 31,61 |
| Rede 108 | Dados de validação (a partir da semana 136) | 108,12 | 10,40 | 41,21 |
| Regressão | Dados de validação (a partir da semana 136) | 189,03 | 13,75 | 49,10 |
| <i>Naive</i> | Dados de criação (até semana 135) | 133,98 | 11,58 | 126,24 |
| <i>Naive</i> | Dados de validação (a partir da semana 136) | 506,44 | 22,50 | 116,09 |

a capacidade de previsão superior da rede neural em relação às outras técnicas.

6. CONSIDERAÇÕES FINAIS

Este trabalho explorou o uso da técnica de redes neurais na previsão de vendas no varejo em nível atomizado.

De acordo com os resultados alcançados na pesquisa, o desempenho das redes foi favorável na previsão de vendas de produtos no nível individual, superando as modelagens *naive* e de regressão linear.

Pode-se concluir que as redes neurais são capazes de tratar os dados de diferentes produtos e que informações agregadas podem auxiliar na previsão de vendas individuais, em razão da existência de relacionamentos entre os produtos. Tais relacionamentos puderam ser captados pelas redes neurais.

A pesquisa mostrou, também, o uso de vendas históricas na previsão, sugerindo o interesse do uso de séries de tempo na previsão de vendas individuais.

Contudo, os erros verificados ainda são altos. A rede neural errou, em termos percentuais, cerca de 41% nas previsões de uma semana à frente — certamente os desvios seriam ainda maiores para horizontes de tempo mais amplos. Ainda há espaço para melhorias significativas na metodologia e técnica de previsão. Fatores como a qualidade dos dados disponíveis, o baixo número de registros por produto, as poucas informações sobre promoções efetuadas, as movimentações macroeconômicas e a falta de informações sobre a concorrência podem explicar o alto erro encontrado nas previsões. Outros estudos poderiam explorar variáveis dessa natureza.

Como foi destacado anteriormente, inconsistências nos dados levaram à eliminação de certo número de registros. Essas inconsistências são inerentes aos processos de garimpagem de dados, que lidam com dados não coletados para esse fim. São gerados a partir das atividades operacionais da empresa, que podem levar a erros de entrada de dados. Ainda assim, as redes neurais demonstraram capacidade de gerar previsões por meio desses dados.

Pode-se destacar, também, que, como a coleta desses dados nos sistemas operacionais não tem por objetivo a implementação de processo de garimpagem de dados, informações sem interesse operacional para a empresa não são sistema-

ticamente coletadas. Dados de ações de propaganda e promoções efetuadas, que possivelmente poderiam explicar variações na demanda de produtos, não são sistematicamente registrados na base de dados, tornando difícil sua obtenção e utilização. Empresas que buscam implementar processos de garimpagem de dados, tal como a empresa estudada, têm interesse em alterar suas rotinas de armazenagem de dados a fim de enriquecer futuras bases de dados voltadas à criação de modelos de previsão. Poderiam, assim, melhorar sua capacidade preditiva de vendas.

Também os dados macroeconômicos poderiam ser testados em pesquisas futuras. Duas hipóteses poderiam ser testadas. Uma primeira hipótese seria assumir que os valores históricos dessas variáveis, isto é, o valor passado de variáveis como crescimento do PIB, variação do dólar, entre outras, geram impacto nas previsões de vendas futuras da empresa. Se, no entanto, assumir-se a hipótese de que o valor futuro dessas variáveis é que impacta no comportamento futuro das vendas, os modelos deveriam incorporar tendências de variáveis econômicas. No entanto, a adoção dessas variáveis exigiria, possivelmente, maior esforço do pesquisador, levando-o também ao estudo de modelos que pudessem prever a tendência de tais variáveis.

O processo de obtenção, seleção e tratamento dos dados mostrou-se bastante trabalhoso e oneroso em termos de processamento dos dados. O grande volume de dados de transações apenas em uma loja fez com que as análises fossem bastante demoradas. Somente para as subamostras selecionadas, o tempo de treinamento das redes neurais levou de três a 12 horas para cada rede. As redes foram treinadas em um microcomputador Pentium. Equipamentos de maior desempenho poderiam aumentar a velocidade de processamento e permitir o treinamento das redes baseado em volume maior de dados.

Finalmente, na pesquisa aqui relatada, optou-se pela aplicação de redes multicamadas e do método de retropropagação proposto por Rumelhart, Hinton e Williams (1986), mais comumente referenciado na literatura. No entanto, certas limitações do método de retropropagação são reportadas, como a tendência a convergir para um ponto ótimo local (STÄGER e AGARWAL, 1997), a falha em utilizar toda a informação estatística relevante (WHITE, 1989), ou mesmo a dificuldade de identificar o ponto ótimo de parada do processo de aprendizado (WARNER e MISRA, 1996). Apesar dessas restrições,

o método é o mais utilizado para treinamento de redes neurais (DETIENNE, DETIENNE e JOSHI, 2003). No entanto, outros modelos e métodos de aprendizado poderiam ser testados,

como as redes de base radial (HAYKIN, 1998), as redes probabilísticas (SPECHT, 1990) ou as redes de regressão generalizada (PATTERSON, 1996; BISHOP, 1995). ♦

NOTA

Dada a dificuldade em reproduzir aqui a imagem gerada pelo *software*, apresenta-se a seguir um dos gráficos gerados a fim de dar ao leitor uma idéia do suporte utilizado para a análise. A intensidade das relações é destacada pela espessura das flechas. Mais detalhes sobre o método podem ser encontrados nos manuais do *software Intelligent Miner* (IBM, 1999).



Relacionamentos entre Produtos Detectados nas Análises do Cesto de Compras

- ACZEL, Amir D. *Complete business statistics*. 2. ed. IL.: Irwin Homewood, 1993.
- ALIBAIG, Shahbaz; LILLY, Bryan. Updating your forecasting system: Wisconsin tissue's experience. *The Journal of Business Forecasting Methods & Systems*, Tucson, v.18, n.3, p.13-18, Fall 1999.
- ALMEIDA, Fernando C.; DUMONTIER, Pascal. O uso de redes neurais em avaliação de risco de inadimplência. *Revista de Administração da Universidade de São Paulo (RAUSP)*, São Paulo, v.31, n.1, p.52-63, jan./mar. 1996.
- ANSUJ, A.P.; CAMARGO, M.E.; RADHARAMANAN, R.; PETRY, D.G. Sales forecasting using time series and neural networks. *Computers and Industrial Engineering*, Nova York, v.31, n.1/2, p.421-425, Oct. 1996.
- BAETS, Walter R.J.; VENUGOPAL, V. Neural networks and statistical techniques in *marketing* research: a conceptual comparison. *Marketing Intelligence & Planning*, Bradford, v.12, n.7, p.30-38, 1994.
- BARASH, Mark; MITCHELL, Donald H. Account based forecasting at Nabisco biscuit company. *Journal of Business Forecasting*, Flushing, v.17, n.2, p.3-6, Summer 1998.
- BASS, Frank M. A new product growth model for consumer durables. *Management Science*, n.15, p.215-227, Jan. 1969.
- BEN-AKIVA, M.; GERSHENFELD, S. Multi-featured products and services: analysing pricing and bundling strategies. *Journal of Forecasting*, Chichester, v.17, n.3-4, p.175-196, June/July 1998.
- BERRY, Michael J.A.; LINOFF, Gordon. *Data mining techniques for marketing, sales and customer support*. 1st ed. [S.l.]: John Wiley & Sons, 1997.
- BISHOP, Christopher M. *Neural networks for pattern recognition*. Oxford: Oxford University Press, 1995.
- BROWN, R. Goodell. *Statistical forecasting for inventory control*. 1.ed. New York: The Maple Press Company, 1959.
- CHURCH, Keith B.; CURRAM, Stephen. Forecasting consumer's expenditure: a comparison between econometric and neural network models. *International Journal of Forecasting*, Amsterdam, v.12, n.2, p.255-267, June 1996.
- DeLURIGO, Stephen A. *Forecasting principles and applications*. New York: McGraw Hill, 1998.
- DETIENNE, Kristen Bell; DETIENNE, David H.; JOSHI, Shirish A. Neural networks as statistical tools for business researchers. *Organizational Research Methods*, Thousand Oaks, v.6, n.2, p.236-265, Apr. 2003.
- FOURT, L.A.; WOODLOCK, J.W. Early prediction of market success of grocery products. *Journal of Marketing*, n.25, p.31-38, Oct. 1960.
- FRANK, Celia; GARG, Ashish; RAHEJA, Amar; SZTANDERA, Les. Forecasting women's apparel sales using mathematical modeling. *International Journal of Clothing Science and Technology*, Bradford, v.15, n.2, p.107-125, 2003.
- GARGANO, Michael L.; RAGGAD, Bel G. Data mining: a powerful information creating tool. *OCLC Systems and Services*, Bradford, v.15, n.2, p.81-90, 1999.
- GORDON, Rich. A role for the forecasting function. *The Journal of Business Forecasting Methods & Systems*, Tucson, v.16, n.4, p.3-7, Winter 1998.
- GROSS, Charles W.; PETERSON, Robin T. *Business forecasting*. 2nd ed. Boston: Houghton Mifflin, 1983.
- GUADAGNI, P.M.; LITTLE, J.D.C. When and what to buy: a nested logit model of coffee purchase. *Journal of Forecasting*, Chichester, v.17, n.3-4, p.303-327, June/July 1998.
- HAIR, J.; BLACK, B.; BABIN, B.; ANDERSON, R.E. *Multivariate data analysis*. 5th ed. Upper Saddle River: Prentice-Hall, 1998.
- HAND, D.J. Data mining: statistics and more? *The American Statistician*, Alexandria, v.52, n.2, p.112-118, May 1998.
- HANSENS, D.M.; PARSONS, J.L.; SCHULTZ, R.L. *Market response models: econometric and time series analysis*. Boston: Kluwer, 1990.
- HAYKIN, Simon. *Neural networks: a comprehensive foundation*. 2nd ed. New York: Macmillan, 1998.
- HIBBERT, Brynn; WILKINSON, Ian F. Chaos theory and the dynamics of *marketing* systems. *Journal of Academy of Marketing Science*, Greenvale, v.22, n.3, p.218-233, Summer 1994.
- HUSSEY, Michael; HOOLEY, Graham. The diffusion of quantitative methods into *marketing* management. *Journal of Marketing Practice: Applied Marketing Science*, Bradford, v.1, n.4, p.13-31, 1995.
- IBM. *Utilizando o Intelligent Miner*, versão 6, release 1. Boeblingen, Germany: IBM Deutschland Entwicklung GmbH Information Development, Dept 0446, Sept. 1999.
- KLEIN, Lawrence R.; YOUNG, Richard M. *An introduction to econometric forecasting and forecasting models*. 2nd ed. Lexington, MA: Lexington Books, 1980. The Wharton Econometric Studies Series.
- KUO, R.J.; XUE, K.C. A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights. *Decision Support Systems*, Amsterdam, v.24, n.2, p.105-126, Dec. 1998.
- LAPEDES, A.; FARBER, R. *Nonlinear signal processing using neural networks prediction and system modeling*. Technical Report LA-UR-87-2662. Los Alamos, National Library, Los Alamos, NM, 1987.
- LAWRENCE, Michael; EDMUNDSON, Bob; O'CONNOR, Marcus. A field study of sales forecasting accuracy and processes. *Amsterdam European Journal of Operational Research*, Amsterdam, v.122, n.1, p.151-160, Apr. 2000.
- LUXHOJ, J.T.; RIIS, J.O.; STENSALLE, B. A hybrid econometric-neural network modeling approach for sales forecasting. *International Journal of Production Economics*, Amsterdam, v.43, n.2-3, p.175-192, 1996.
- MAHAJAN, V.; BASS, Frank M.; MULLER, E. New product diffusion models in marketing: a review and directions for

- research. *Journal of Marketing*, v.54, p.1-26, Jan. 1990.
- MAKRIDAKIS, S.; WHEELWRIGHT, Steven C.; MCGEE, Victor E. *Forecasting: methods and application*. 2nd ed. New York: John Wiley & Sons, 1983.
- MANSFIELD, E. Technological change and the rate of imitation. *Econometrica*, v.29, n.10, p.741-766, 1961.
- MASON, J.B.; MAYER, M.L.; WILKINSON, J.B. *Modern retailing: theory and practice*. 6th ed. Chicago: Irwin, 1993.
- MENTZER, John T.; KENT, John L. Forecasting demand in the Longaberger Company. *Marketing Management*, Chicago, v.8, n.2, p.46-50, Summer 1999.
- METAXIOTIS, Kostas; PSARRAS, John. The contribution of neural networks and genetic algorithms to business decision support. *Management Decision*, Londres, v.42, n.2, p.229-242, 2004.
- PARENTE, Juracy. *Varejo no Brasil: gestão e estratégia*. São Paulo: Atlas, 2000.
- PATTERSON, Dan. *Artificial neural networks*. Singapore: Prentice Hall, 1996.
- ROBERTS, John. *Marketing approaches to forecasting problems*. *Journal of Forecasting*, Chichester, v.17, n.3-4, p.169-174, June/July 1998.
- RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning internal representations by error propagation. In: RUMELHART, D.E.; MCCLELLAND, J.L. (Ed.). *Parallel distributed processing*. London: MIT Press, 1986. v.1.
- RUMELHART, David E.; MCCLELLAND, James C. PDP Research Group (Parallel Distributed Processing) — exploration in the microtexture of cognition. London: The MIT Press, 1986. v.1.
- SPECHT, Donald F. Probabilistic neural networks. *Neural Networks*, Oxford, v.3, n.1, p.109-118, 1990.
- SPEDDING, T.A.; CHAN, K.K. Forecasting demand and inventory management using Bayesian time series. *Integrated Manufacturing Systems*, v.11, n.5, p.331-339, 2000.
- STÄGER, F.; AGARWAL, M. Three methods to speed up the training of feedforward and feedback perceptrons. *Neural networks*, Oxford, v.10, n. 9, p.1435-1443, Nov. 1997.
- SUBRAHMANYAN, Saroja. Using quantitative models for setting retail prices. *Journal of Product and Brand Management*, Santa Barbara, v.9, n.5, p.304-320, 2000.
- TANG, Z.; ALMEIDA, C.; FISHWICK, P.A. Time series forecasting using neural networks vs. box-jenkins methodology. *Simulation Council*, v.57, n.5, p.303-310, Nov. 1991.
- THALL, N. Neural forecasts: a retail sales booster. *Discount Merchandiser*, v.32, n.10, p.41-42, 1992.
- TUKEY, J.W.; HOAGLIN, D.C.; MOSTELLER, F. *Análise exploratória de dados: técnicas robustas*. Lisboa: Salamandra, 1983.
- WARNER, B.; MISRA, M. Understanding neural networks as statistical tools. *American Statistician*, Alexandria, v.50, n.4, p.284-293, Nov. 1996.
- WHITE, H. Neural network learning and statistics. *AI Expert*, v.3, n.4, p.48-52, Dec. 1989.
- WRAY, Barry; BEJOU, David; PALMER, Adrian. Using neural network analysis to evaluate buyer-seller relationships. *European Journal of Marketing*, Bradford, v.28, n.10, p.32-48, 1994.
- ZHANG, Guoqiang; PATUWO, B. Eddy; HU, Michael Y. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, Amsterdam, v.14, p.35-62, 1998.

Using neural networks for sales forecast in retail industry

This paper explores the use of artificial neural networks in sales forecasts in retailing industry. Historical data is used to make forecasts of product sales in the retailing industry. Weekly sales data from individual products were used to generate models. Results suggest that models based upon neural networks can be adequate to individual product sales forecast. Models obtained with the proposed methodology are able to predict product sales in the short term more accurately than non adjusted *naïve* techniques and linear regression, more frequently used.

Uniterms: sales forecast, retailing industry, artificial neural networks.

Previsión de ventas en el comercio minorista por medio de redes neurales

En este trabajo, se investiga el uso de redes neurales en la previsión de ventas al por menor. Con esa técnica se elaboraron previsiones a partir de datos históricos de ventas de productos de una empresa minorista. Se generaron modelos con el uso de datos de venta de cada uno de los productos de la base de datos, semana a semana. Los resultados sugieren que el modelado por medio de redes neurales artificiales puede ser considerado adecuado para la previsión de demanda de productos considerados individualmente (producto por producto). Los modelos obtenidos con el uso de la metodología propuesta pueden prever las ventas de productos a corto plazo con mayor precisión que las técnicas *naïve* no ajustada y de regresión lineal, más frecuentemente utilizadas.

Palabras clave: previsión de ventas, comercio minorista, redes neurales.