Solução Baseada em Redes Bayesianas para Prevenção de Fraudes em Cartões de Crédito

Carla Barata Ribeiro[2], Cleuber Moreira Fernandes[2], Fernando Henrique de Souza Santos[1], Ronaldo Câmara de Araújo[2], Tereza Cristina da Costa[1]

[1]Centro Tecnológico do Banco do Brasil
STN 716 Conjunto C Ed. Sede IV – 70.770-100 – Brasília - DF - Brasil
{teres,fernandoh}@bb.com.br

[2]Politec Informática Ltda
SIG quadra 4 lote 173 – 70.610-440 – Brasília – DF – Brasil
{cleuber.fernandes,ronaldo.camara,carla.barata}@politec.com.br

Abstract. *This paper presents the experience of the Nucleus of Artificial Intelligence of the Banco do Brasil in the development of a solution for the problem of prevention of frauds in credit card through a tools set combined to subsidize business analysts' decision. The solution uses Bayesianas Networks Models and architecture multi-platform.*

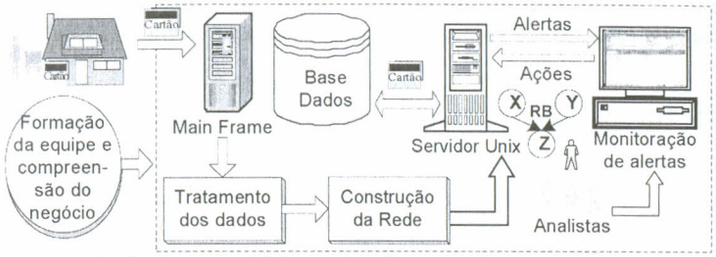
Resumo. *Apresenta-se a experiência do Núcleo de Inteligência Artificial do Banco do Brasil no desenvolvimento de uma solução para o problema de prevenção de fraudes em cartão de crédito através de um conjunto de ferramentas que são combinadas de forma a subsidiar a tomada de decisão dos analistas de negócio. A solução utiliza modelo de redes Bayesianas e arquitetura multiplataforma.*

1. Introdução

O Banco do Brasil atua no mercado financeiro, captando e aplicando recursos de terceiros. A grande diversificação de produtos gera crescimento perceptível no número de transações financeiras realizadas por meio de cartões de crédito. Dessa forma, a Empresa tem a necessidade de criar mecanismos que possam atuar na proteção das transações bancárias dos clientes, para diminuir o risco de imagem e aumentar a perda evitada. Fica assim evidenciada a necessidade de se desenvolver uma solução que reduza efetivamente este prejuízo. No entanto, a probabilidade a priori de fraude é muito pequena ($5x10^{-4}$) e as características de fraude são diversificadas, o que torna consideravelmente complexa a compreensão do fenômeno fraude e a construção de um modelo que o descreva.

Este artigo apresenta o estudo de caso de aplicação de um sistema especialista utilizando Rede Bayesiana (RB), técnica de Inteligência Artificial (IA) efetiva no suporte à decisão sob condições de incerteza, como solução para prevenção de fraudes em cartões de crédito. Uma Rede Bayesiana permite a compreensão do fenômeno por sua representação gráfica e probabilística, e o aproveitamento do conhecimento empírico dos especialistas do domínio.

No processo de concepção da solução, verificou-se a importância de uma abordagem que consolidasse não somente técnicas e ferramentas de IA, mas também estabelecesse um tratamento completo, desde a identificação de transações com alta probabilidade de fraude até a efetiva ação dos analistas de negócio envolvidos no



monitoramento do fenômeno. A Figura 1 exibe sumariamente a arquitetura da solução composta por quatro principais fases combinadas num ambiente multiplataforma.

Figura 1. Arquitetura da solução NIA

Inicialmente, é necessário formar uma equipe envolvendo especialistas da área de negócio, de IA e analistas de sistemas, bem como compreender conjuntamente as peculiaridades da área de negócio (seção 2). Em seguida, deve-se proceder o tratamento dos dados, buscando identificar e categorizar as variáveis mais significativas (seção 3). Posteriormente, o processo de construção e avaliação da Rede Bayesiana é executado para gerar um modelo que represente o fenômeno (seção 4). Por fim, os alertas de fraude produzidos pela RB serão monitorados pelos analistas de negócio a partir de uma ferramenta Web (seção 5).

2. Formação da Equipe e Compreensão do Negócio

No princípio, a equipe de IA propõe uma apresentação da solução aos interessados (áreas de negócio), com o objetivo de mostrar a tecnologia a ser adotada, funcionamento, e benefícios. É necessário salientar a importância do envolvimento da área de negócio no entendimento, acompanhamento da solução e a participação nos riscos eminentes, tais como: dados inconsistentes que não representam o domínio, ambiente não suportar o processo de inferência e falta de envolvimento do especialista.

A equipe do Núcleo de Inteligência Artificial (NIA) atua na coordenação inicial do projeto, apresentando processos necessários e respectivas formas de atuação, de acordo com a figura 2.

Figura 2. Formação da equipe

Nesse momento, é possível proceder a formação da equipe, de acordo com o perfil desejado e disponibilidade necessária. Os participantes dessa cadeia devem ter, de forma estruturada e atualizada, a descrição dos processos, de modo a permitir que todos saibam exatamente o trabalho a ser feito, quem deverá executá-lo, quando e onde. Para tanto, é gerado um documento de iniciação do projeto, contendo os objetivos, responsável e demais participantes envolvidos; restrições e cronograma preliminar. Após aprovação desse documento, começa o trabalho em conjunto com a equipe.

A conclusão dessa fase ocorre com a compreensão do negócio, onde a área de negócio organiza apresentações de: características do problema, das próprias necessidade e relacionamento com outros sistemas.

Na etapa de compreensão do negócio, é possível identificar os fatores que poderão influenciar os resultados positivamente ou negativamente.

3. Tratamento dos Dados

Nesta fase da solução, foi adotada a metodologia CRISP-DM [CRISP 1996] fundamentada em técnicas de *Data Mining*, que norteou o processo de compreensão e preparação dos dados que são utilizados para gerar os modelos de RB (seção 4).

3.1 – Compreensão e Preparação dos Dados

A preparação e compreensão dos dados consistem em identificar todas as informações relevantes para o modelo. É o momento de compreender seu conteúdo, descrição, qualidade e utilidade. Para tanto, emprega-se tratamento estatístico e sugestões do especialista.

É imprescindível a elaboração de variáveis informativas a partir de um subconjunto de dados brutos contidos numa base de dados, como exemplo, a freqüência de compras num determinado estabelecimento.

Trabalhar com grandes massas de dados constitui uma tarefa onerosa, às vezes, impraticável. Neste trabalho, os dados reais de transações foram amostrados várias vezes, com diferentes vieses, na tentativa de gerar amostras com maior probabilidade a priori de fraude. Esta heurística é necessária para que os algoritmos de aprendizagem de RB encontrem um número maior de dependência entre a variável alvo (fraude) e as demais variáveis informativas para o fenômeno. As amostras foram geradas com $5x10^{-2}$ de transações de clientes que nunca tiveram fraude, mais todas as transações que constituem fraude no período proposto. Este viés produziu resultados satisfatórios (Seção 4).

3.2 – Seleção de Variáveis

O conjunto de variáveis criadas para se fazer o estudo pode ser Qualitativo (Nominal, Ordinal, Escalar e Razão) e Quantitativo (Contínuas e Discretas) sendo este grupo a grande maioria delas. A variável endógena, ou seja, de interesse a se estimar é do tipo dicotômica, fraude ou não fraude.

Inicialmente, foi realizada uma análise descritiva para se estudar o comportamento das variáveis, observando sua distribuição sempre levando em conta a variável alvo fraude. Para essas análises foram utilizadas representações gráficas; medidas descritivas; posição e variação, de modo a tentar compreendê-las. Observou-se também seus *outliers*, sua concentração, sua distribuição de modo a explicar, de forma mais contundente o fenômeno fraude.

Após essa análise inicial, foram utilizadas técnicas de análise multivariada de forma a identificar quais variáveis conseguem explicar o fenômeno de interesse. Essas

variáveis podem ser combinadas de modo a se criar outras que consigam, com um número bem menor de variáveis, explicar as demais, onde se destaca a técnica de Componentes Principais. Observou-se a interdependência entre as variáveis, das quais se destacam as medidas de covariância e a correlação. Foram selecionadas aquelas que possuem menores medidas, partindo do pressuposto de que se a dependência é alta, uma ou outra variável poderá explicar o fenômeno fraude de forma aproximadamente igual ao modelo. Entre as técnicas de modelagem, destacou-se a Regressão Logística, uma vez que a variável endógena, isto é, a que almeja-se estimar é do tipo binário.

Nessa análise, foram utilizados processos de seleção de variáveis, entre eles destaca-se o método *stepwise*, onde se faz todas as combinações possíveis de inclusão das variáveis, calcula-se a estatística do teste, a saber, Qui-Quadrado de Wald e verifica-se se essa variável é significativa, esse processo é interativo de modo a ficar com o modelo que possua as variáveis mais significativas.

Ao fazer essa seleção, observou-se, pelo grande número de variáveis, a necessidade de se eliminar mais variáveis. Nesse caso, foram utilizados os pesos dos parâmetros estimados (β_k), Qui-Quadrado de Wald, o *ODDS Ratio*, que têm a finalidade de quantificar a chance que o evento de interesse tem de ocorrer.

Fez-se necessário, nesse estudo, uma análise de multicolinearidade entre as variáveis, isto é, observar ainda se entre as variáveis endógenas existem fortes dependências lineares. Utilizada essa técnica para selecionar variáveis, foi realizada uma análise multivariada de *Cluster* de variáveis, que consiste em investigar a correlação entre um conjunto de variáveis, de modo a agrupar as variáveis que possuem características bastante semelhantes.

Outra técnica utilizada para corroborar os resultados obtidos pelas técnicas acima, é a árvore de decisão, que utiliza testes de interdependências χ^2 de forma a segmentar a variável por ordem de importância para a variável de interesse. Por exemplo, a variável que se liga diretamente à variável alvo, nesse caso fraude é a que tem maior dependência, observa-se em qual ramo obtém-se maiores probabilidades de deflagrar fraude, seleciona-se a segunda, e assim sucessivamente de forma a obter as que possuem maiores importâncias.

3.2 – Categorização de Valores

Uma vez selecionadas as variáveis que melhor explicam o fenômeno fraude, há a necessidade, no caso de variáveis quantitativas, discretas ou contínuas, de se agrupar esses valores em faixas. Isto ocorre porque os algoritmos de aprendizagem de redes Bayesianas, implementados nessa aplicação, utilizam o conceito de tabelas de contingência de modo a representar as probabilidades condicionadas das famílias, isto é, uma variável e seus pais.

Nesse passo, precisa-se também da análise descritiva das variáveis. Essa análise utiliza diversas medidas de posição, tais como média, mediana, moda, percentis bem como medidas de variabilidade: variância, desvio padrão, coeficiente de variação, amplitude total, amplitude interquatílica e para conhecer o comportamento com relação à disposição dessas variáveis, utilizou-se também medidas de assimetria e curtose. Outra técnica bastante aliada a esse passo é a árvore de decisão, e dentre os algoritmos utilizados, destaca-se o de *CHAID - Chi-square Automatic Interaction Detectors*, que entre outras funções agrupa valores em faixas de forma a maximizar a dependência entre as variáveis.

Evidencia-se que, em todo o processo, é de fundamental importância a presença do especialista do domínio, que pode tanto incluir como excluir variáveis e sugerir faixas de valores, que com o acúmulo de sua experiência contribui para o entendimento das variáveis que melhor explicam o fenômeno da fraude.

4. Construção da Rede Bayesiana

Uma Rede Bayesiana (RB) é um modelo, que combina as principais características da Teoria dos Grafos e Teoria da Probabilidade, apropriado para representar conhecimento e dar suporte à tomada de decisão sob condição de incerteza. Uma RB é constituída de um Grafo Orientado Acíclico e uma tabela de probabilidades condicionais associada a cada variável. Pode-se construir uma RB a partir do conhecimento de um especialista ou automaticamente a partir de dados representativos do domínio [Fernandes 2004].

Na solução apresentada, a construção da RB é realizada automaticamente a partir de dados, sendo constituída de quatro etapas (Figura 3). Foram implementados os algoritmos K2 e B utilizando as métricas Bayesiana Padrão e *Cooper and Herskovits* [Castillo et al. 1997] e o algoritmo TPDA utilizando a métrica Informação Mútua (IM) [Cheng et al. 2003]. Estes podem ser utilizados na etapa de treinamento da rede usando uma amostra de dados de oito meses de transações, enviesada em 1,07% de fraude. Este viés é necessário para reforçar a identificação de relações de dependência entre as variáveis, uma vez que a probabilidade a priori de fraude é muito pequena $5x10^{-3}$%. A etapa de pontuação usa uma implementação do algoritmo de inferência baseado em Árvore de Junção [Ladeira et al. 1999]. Essa etapa infere uma probabilidade para a variável alvo, a partir de uma RB e de uma amostra de dados de um mês de transações, utilizada para validar os algoritmos de treinamento citados acima. A aferição usa o arquivo pontuado e um limiar da priori para produzir a matriz de confusão, utilizada para avaliar a qualidade da rede.

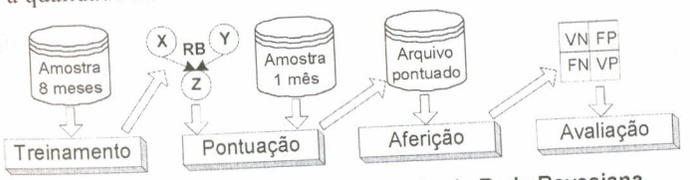


Figura 3. Etapas da fase de construção da Rede Bayesiana

A matriz de confusão provém dos tipos de erro I e II do teste de hipótese estatístico. A hipótese nula é se uma determinada transação é fraude. O falso positivo (FP) corresponde ao tipo de erro I, enquanto que o falso negativo (FN) corresponde ao tipo de erro II. O verdadeiro positivo (VP) é o complementar do FN, enquanto que o verdadeiro negativo (VN) é o complementar de FP. Objetiva-se maximizar o VP e VN, e minimizar o FP e FN, ou seja, produzir o maior número de alertas de fraudes sobre transações que realmente configuram fraudes.

Os gráficos na Figura 4 apresentam os resultados obtidos em dois experimentos, utilizando amostras com diferentes conjuntos de variáveis e categorias de valores, denominados rede 10 e 19. Ambas as amostras passaram pelo processo de construção de

RB (Figura 3). Observou-se que a variação dos resultados é muito pequena. Os algoritmos K2 e TPDA produziram aproximadamente os mesmos números de acertos e erros considerando a rede 10, mas o algoritmo B conseguiu identificar mais fraudes (VP), apesar de incorrer em maior número de alertas falsos (FP). Os três algoritmos produziram melhores resultados com a rede 19. No entanto, o algoritmo B ainda foi melhor que os demais, conseguindo identificar 80% das transações fraudulentas (VP) e reduzindo os alertas falsos (FP) para apenas 14% do volume de transações.

Por limitação na capacidade de tratamento de alertas, aproximadamente 400 alertas ao dia, objetiva-se a redução do FP para 5% do volume de transações. Estrategicamente, espera-se alcançar essa redução submetendo aquelas transações que tenham sido consideradas alertas de fraude a uma segunda rede que contenha informações sobre o perfil de compra dos clientes.

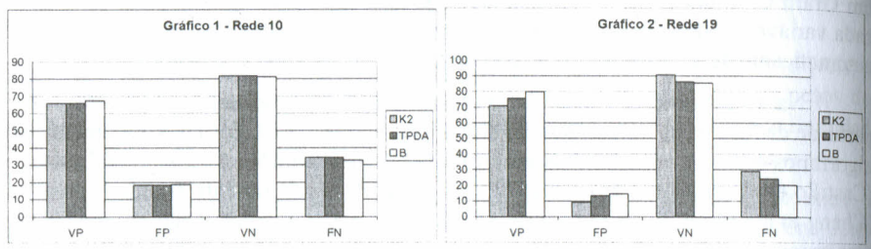


Figura 4. Resultados produzidos pelos algoritmos K2, B e TPDA

O algoritmo K2 exige que uma ordenação causal das variáveis seja informada, para proceder à orientação do grafo. Contudo, não se conhece essa ordenação, o que pode levar a orientações incorretas prejudicando os resultados. O algoritmo TPDA utiliza a métrica IM e um limiar ε que influência o número de arestas inseridas no grafo, o que também pode comprometer os resultados. Os melhores resultados produzidos pelo algoritmo B pode ser explicado pela inexistência de limitações, como as existentes nos demais algoritmos. O impacto dessas limitações pode ser minimizado com obtenção de dados que melhor representem o fenômeno fraude, com variáveis informativas e amostras com suficiência estatística.

5. Ferramenta de Monitoração de Alertas

Uma vez gerados os alertas (transações com probabilidade de fraude) por meio de uma Rede Bayesiana, é preciso apresentá-los para que possam ser interpretados (quanto ao risco de fraude) e para que ações, como bloquear um cartão, possam ser tomadas por analistas do negócio em questão. É importante para isso que uma ferramenta de monitoração seja construída para a execução deste tratamento. Além disso, uma ferramenta de monitoração de alertas eficiente permite manter o modelo gerado por meio da Rede Bayesiana atualizado quanto às variações do fenômeno da fraude. A ferramenta desenvolvida é composta por três módulos, como segue.

5.1 Módulo para tratamento dos alertas

Apresenta as probabilidades inferidas pela Rede Bayesiana, associadas a cada elemento avaliado, que no caso apresentado são transações feitas com cartão de crédito. Nesta visualização, os dados importantes para a análise de negócio devem ser também apresentados. Além disso, a linguagem adotada deve ser a utilizada usualmente pelos analistas de negócio, para que os mesmos se sintam familiarizados com a ferramenta.

Permite que os analistas trabalhem em conjunto, garantindo o máximo de produtividade possível. O princípio utilizado foi o da fila de atendimento, se um analista de negócio está tratando as transações de um certo cliente, outro analista não pode mais visualizar as transações do mesmo cliente, garantindo assim a integridade do processo.

Permite que os analistas selecionem as características com as quais desejam obter as transações que irão tratar. O analista pode visualizar um conjunto de alertas para escolher o cartão que irá tratar (atuando nos alertas de mais altas probabilidades), pode estabelecer condições e receber o cartão com a transação mais pontuada dentro das características informadas (atuando por meio de um filtro) ou por meio do número do cartão específico que o mesmo deseja analisar (atuando diretamente em um cartão).

Para maior produtividade no tratamento dos alertas, é necessário disponibilizar outras informações, tais como dados da fatura, não diretamente para a análise dos alertas, mas para o contato com os clientes, como é o caso da monitoração de transações em cartão de crédito. Estas devem estar acessíveis através da ferramenta de monitoração, caso contrário, o custo para buscá-las em outro sistema pode ser prejudicial.

5.2 Módulo para gestão dos analistas e do fenômeno de fraude

A ferramenta permite a gestão do trabalho dos analistas envolvidos na monitoração. Essa facilidade é importante para que se possa verificar o uso da ferramenta, se os alertas estão sendo tratados na íntegra (ou se estão em maior ou menor quantidade do que é possível tratar), e se as ações dadas estão sendo eficazes. Um analista, com perfil de supervisor, pode acompanhá-las. Ações de descarte podem ser acompanhadas individualmente. Além disso, verifica-se a quantidade de transações com certas características, como as provenientes de determinada unidade da Federação entre outras, para melhor direcionar o trabalho dos analistas.

5.3 Módulo para gestão da rede

A ferramenta oferece estatísticas que mostram qual o índice de acerto da Rede Bayesiana para que seja avaliada a necessidade de ajustes na mesma quando os índices obtidos não estiverem satisfatórios. Analisando o FP e o FN, pode-se decidir pela geração de um novo modelo (construção de uma nova Rede Bayesiana).

Características secundárias, tais como a agregação de alertas vindos de outras ferramentas de I.A. ou de outros meios de análise, como regras de negócio fixas que possam ser aplicadas, também são interessantes em uma ferramenta de monitoração eficiente.

Optou-se por uma aplicação de monitorização voltada à *Intranet*. A implementação foi feita em linguagem de programação Sun Java, principalmente por sua portabilidade, usando-se como metodologia a construção em três camadas (Modelo, Visão e Controle). Além disso, devido às características de acesso às bases de dados encontradas no ambiente da aplicação, modelaram-se classes para enviar e receber dados das mesmas, que serviram como interface com programas Natural (que de fato realizam quaisquer consultas). A identidade visual da aplicação bem como a navegação

pela mesma seguiram as normas corporativas e as demandas dos usuários finais, e fora orientadas pelos conceitos já estabelecidos de desenho de aplicativos para redes d computadores.

6. Conclusões e Trabalhos Futuros

Este trabalho contribui com uma visão macro de cada fase essencial para desenvolvimento e implantação de sistemas especialistas apropriados para solucion problemas de decisão incerta. Percebe-se que pesquisadores acadêmicos da área de I dominam conceitos e técnicas apropriadas para aplicações com o escopo aqu apresentado. No entanto, a experiência prática é indispensável para consolidar solução. Pode-se destacar como contribuição empírica a utilização de amostra enviesadas, a criação de variáveis informativas a partir de dados brutos e a necessidade de tratamento estatístico dos dados a serem utilizados na geração dos modelos de Redes Bayesianas.

A solução foi implantada e homologada pela área de negócio de cartões de crédito. Os resultados obtidos até o presente momento são satisfatórios, mas não suficientes. Demanda-se, ainda, esforço na tentativa de produzir modelos que tenham uma taxa de sucesso maior. Espera-se obter melhores resultados criando um sistema multicamadas de RB especializadas. Também será implementado um processo que leve em consideração a perda evitada, isto é, direcionar os esforços de tratamento daquelas transações com maior probabilidade de serem fraude e também com maiores valores monetários. A aplicação desta solução será expandida a outras áreas de negócio, tais como análise de crédito e lavagem de dinheiro. A pesquisa continua, objetivando aumentar a acurácia da ferramenta. Novos resultados serão publicados futuramente.

Referências

Allison, P. D. (2001) "Logistic Regression Using The SAS System: Theory and Application", Edited by SAS Institute Inc. and John Wiley & Sons ltd.

Araújo, Ronaldo C. (1998) "Controle Estatístico de Processos Multivariados". Dissertação de Mestrado. Universidade de Brasília, Brasília, DF.

Castillo, E., Gutiérrez, J. M. and Hadi, A. S. (1997) "Expert systems and probabilistic network models". Monographs in Computer Science, Springer-Verlag, New York.

Cheng, J. et al. (2002) "Learning belief networks from data: An information-theory approach". The Artificial Intelligence Journal, Vol 137, p. 43-90.

CRISP (1996) "Cross Industry Standard Process for Data Mining". http://www.crisp-dm.org/Process/index.htm acessado em 19/08/2004.

Fernandes, Cleuber. M. da Silva, W. T., Ladeira, M. (2004) "An Algorithm to Learning Bayesian Networks from Small Datasets". In: Simpósio de Informática da Região Sul, 3. Santa Maria, RS.

Ladeira, M.; Viccari, R. M.; Coelho, H. (1999) "Raciocínio Probabilístico em Sistemas Inteligentes". In: Congresso da Sociedade Brasileira de Computação, JAI.