

Análise do perfil do usuário de serviços de telefonia utilizando técnicas de mineração de dados

Adelir José Schuler Junior¹, Anderson Luiz Fernandes Perez^{1,2},

¹Faculdades Barddal
Curso de Sistemas de Informação
Trindade - 88036-500 – Florianópolis – SC
adelir@barddal.br

²Universidade Federal de Santa Catarina (UFSC)
Departamento de Automação e Sistemas
Curso de Pós-Graduação em Engenharia Elétrica
Campus Universitário – Trindade - 88040-900 – Florianópolis – SC
anderson@das.ufsc.br

Resumo

O prejuízo causado às operadoras de telecomunicações pela inadimplência cresce assustadoramente. Na tentativa de minimizar esse prejuízo, o setor vem adotando algumas medidas. As técnicas de mineração de dados para a descoberta de conhecimento trazem contribuições importantes para o setor de telecomunicações. A mineração de dados apresenta-se como um recurso poderoso na identificação de perfis de usuários dos serviços de telecomunicações, fornecendo subsídios no combate a inadimplência, assim, pode-se apresentar informações que justificam a montagem de uma estratégia e uma estrutura adequada para combater a grande quantidade de fraudadores. Quanto mais rápido forem identificados estes usuários, menor será o prejuízo da operadora de telefonia e mais recursos a empresa terá para oferecer novos serviços. Este trabalho propõe o emprego de técnicas de *Data Mining* para analisar e identificar o perfil do usuário inadimplente no setor de telecomunicações.

Palavras-chave: Telecomunicação, Mineração de Dados.

Abstract

The loss of income due to non-payment is increasing in an amazing rate to the Telecommunication Companies. In an attempt to minimize this loss, companies are taking some measures, like, for instance, *Data Mining*. Data Mining is a way of "mining", or reseraching, data from customers aiming to discover traits that could help to identify bad payers. It can be a powerful resource in the profiling of those customers, giving the companies important information in order to attack the origin of the loss. With Data Mining we are able to present information that justify the development of an adequate strategy/structure to combat this legion of cons. The faster these users are identified, the lesser the loss and the bigger the resources the Company will have to invest in new services. Data Mining suggests different tecniques to maximise the data discovery, and to get the most of Data Mining you need to be familiarized with them. The aim of this work is to compare this Data Mining tecniques, which are used in the profiling of the users, with an eye on the non-payers, in our Telecommunication sector. In some cases, a combination of tecniques may be used to get the solution.

Key-words: Telecommunication, Data Mining.

1 Introdução

As operadoras de telecomunicações tem absorvido prejuízos consideráveis causados pela inadimplência ou pela utilização de seus recursos de forma ilícita.

As técnicas de mineração de dados aplicadas no setor de telecomunicações apresentam-se como um recurso poderoso para identificar o perfil do inadimplente, fornecendo subsídios para as operadoras de telecomunicações no combate a este problema, assim, é possível apresentar dados que dão base sólida para

montar uma estrutura adequada para combater o grande número de fraudadores. Quanto mais rápido forem identificados estes usuários, menor será o prejuízo da operadora de telefonia e mais recursos à empresa terá para oferecer novos benefícios aos usuários.

Além da determinação do perfil dos usuários inadimplentes, técnicas de mineração de dados podem ser utilizadas para apresentar dados sobre o perfil dos bons pagadores mostrando qual a preferência na utilização dos serviços, dando embasamento para que as empresas do setor de telecomunicações tenham condição de realizar campanhas de marketing para manter e agregar bons clientes.

O objetivo deste trabalho é apresentar o uso das técnicas de mineração de dados como: redes neurais (RNAs) e árvores de decisão, a fim de avaliar qual dessas técnicas se torna mais eficaz para identificar o perfil de um usuário inadimplente.

As técnicas de mineração de dados (redes neurais e árvores de decisão) são aplicadas sobre os dados que as operadoras de telefonia dispõe referente a usuários inadimplentes. A utilização dessas técnicas busca determinar o perfil (classe) do usuário inadimplente e também criar regras para testar novos dados verificando se encaixam nos perfis (classes) de inadimplência. Desta forma pode-se instituir alarmes informativos permitindo a operadora de telefonia tomar providências imediatas, evitando que a descoberta da fraude aconteça somente muito tempo depois da utilização do recurso (utilização do telefone), diminuindo os prejuízos causados.

2 O problema que a inadimplência causa no setor de telecomunicações

Passado algum tempo das privatizações no setor de telecomunicações, as operadoras estão de frente para um novo problema, o crescimento diário da inadimplência, que segundo algumas estimativas superam a marca de R\$ 2 bilhões [4].

Vários motivos são atribuídos a esta alta taxa de inadimplência: alta taxa de ICMS, tarifas altas, recessão econômica, o desemprego e conseqüentemente a queda do poder aquisitivo da população e as fraudes causadas pelo crime organizado, são alguns dos fatores determinantes para a inadimplência. O problema da inadimplência está diretamente relacionado ao problema econômico do país **Error! Reference source not found.** Diante deste cenário, pode-se perceber que as despesas das operadoras de telefonia com a inadimplência são grandes, cerca de 10% da receita bruta **Error! Reference source not found.**

Na tentativa de minimizar estes custos, a ANATEL tenta ajudar as operadoras com algumas regras [5]:

- a) nos primeiros 15 dias: a prestadora notificará ao assinante, dos seus direitos de contestação do débito, em até 5 dias úteis, e da possibilidade de suspensão parcial do serviço por inadimplência;
- b) a partir de 30 dias de inadimplência: suspensão parcial do serviço. A suspensão será apenas para fazer ligações e não para recebê-las;
- c) após 30 dias da suspensão parcial: permanecendo o assinante inadimplente, poderá haver suspensão total. Neste caso não poderá ser feita ligações e nem recebê-las;
- d) deverá ser comunicado ao assinante, com antecedência mínima de 15 dias, a suspensão total do serviço;
- e) transcorridos 30 dias de suspensão total do serviço em determinada modalidade do STFC, por inadimplência, a Prestadora pode rescindir o contrato de prestação de serviço;
- f) rescindido o contrato de prestação de serviço, por inadimplência, a prestadora pode incluir o registro de débito em sistemas de proteção ao crédito;
- g) caso o assinante conteste os débitos, os prazos estabelecidos estarão suspensos até que o mesmo seja notificado da decisão da Prestadora;
- h) na hipótese de contestação parcial, deve ser suspensa a cobrança da parcela impugnada e efetuado o pagamento da parte incontroversa;
- i) é direito do assinante, sem qualquer custo, requerer a retificação das informações de inadimplência relativas a sua pessoa;
- j) a prestadora responde pelos danos de qualquer natureza em razão de informação inverídica de inadimplência; e
- k) assinante inadimplente pode efetuar a qualquer momento o pagamento do débito, acrescido dos encargos de mora, cabendo a prestadora retirar a informação de inadimplência e restabelecer o serviço em até 24 horas após a declaração ou comprovação do pagamento pelo assinante ou, ainda, do recebimento pela prestadora de comprovação do pagamento via sistema bancário, se não houver sido rescindido o contrato de prestação de serviço.

Mesmo, com esse conjunto de regras, a taxa de inadimplência continua alta, e para tentar diminuir esta taxa as operadoras de telefonia vêm investindo em novos recursos tecnológicos.

O que muitas empresas não percebem, é que buscar uma solução para o problema da inadimplência pode estar mais próximo do que o imaginado. As informações necessárias para entender as fraudes está no banco de dados. Ignorando isso as empresas buscam sempre soluções perfeitas para seus problemas utilizando o *Data Warehouse* somente para gerar

relatórios não trazendo desta forma muitos benefícios para a corporação **Error! Reference source not found.**

3 Mineração de Dados em uma Base de Dados de Chamadas Telefônicas

Descobrir conhecimento em um banco de dados através da mineração de dados requer organização e capacidade de escolher as características relevantes que serão utilizadas para que o aprendizado possa acontecer.

Apresenta-se a aplicação das técnicas: árvores de decisão, que decompõe um grande problema em subproblemas mais simplificados tornando a análise mais fácil, e redes neurais, que tem a capacidade de “aprender” a executar tarefas por meio de treinamentos. Ambas as técnicas são aplicadas sobre um volume de dados de testes para obtenção de resultados que demonstrem o funcionamento da mineração de dados na busca de conhecimento.

Com o passar do tempo o volume de dados gerados dentro das empresas aumenta. Os locais de armazenagem são modificados para suportar essa quantidade de dados, as tecnologias de controle de armazenagem são melhoradas, enfim, cada dia fica mais fácil armazenar grandes volumes de dados e depois recuperá-los. A grande indagação que se tem nos dias de hoje é: será que os dados armazenados podem ser utilizados com outros fins, ou seja, será que é possível aproveitar esses dados para descobrir se existe algum conhecimento escondido? Por exemplo, uma empresa prestadora de serviços em telecomunicações pode ter em seus registros informações que apresentem tendências de mercado ou mesmo informações que revelam o perfil do seu cliente.

Um supermercado pode apresentar perfis de usuários que mostrem as preferências de consumo, baseado nisso é possível dispor as mercadorias para que fique mais fácil e confortável a compra. Na área da medicina pode-se realizar a busca de informações para diagnosticar certas doenças **Error! Reference source not found.**

Isto já é justificativa suficiente para os esforços aplicados na busca deste conhecimento escondido que um grande volume de dados pode ter. A busca destas informações pode ser feita utilizando alguns técnicas diversificadas, o que se fala muitos nos dias de hoje é a técnica de *Data Mining* **Error! Reference source not found.**

Data Mining ou mineração de dados, é o processo de descobrir informações relevantes, como padrões, associações, mudanças, anomalias e estruturas, em grandes quantidades de dados armazenados em banco de dados, depósitos de dados ou outros

repositórios de informação **Error! Reference source not found.**

3.1 Formulação do problema

O conhecimento do perfil dos inadimplentes, ou o mapeamento da situação da inadimplência no setor de telecomunicações é alvo de estudos e pesquisas constantes. Há necessidade de respaldar as operadoras de telecomunicações com informações que possam melhorar a confiabilidade e a qualidade das informações sobre os usuários inadimplentes.

Utilizando-se o registro das chamadas telefônicas dos usuários inadimplentes que as operadoras possuem, pode-se descobrir e montar regras de classificação que permitem a definição do perfil dos usuários inadimplentes no setor de telecomunicações. Com as regras de classificação definidas, pode-se realizar a detecção de prováveis futuros inadimplentes com a aplicação da regra de classificação sobre uma nova chamada registrada. Caso a chamada encaixe-se em uma das regras definidas, pode-se atribuir valores acumulando pontuação até que tenha-se um índice para especificar se o usuário tem o perfil de inadimplente.

3.2 Dados utilizados para a realização dos testes

Os dados das chamadas telefônicas utilizados para a realização dos experimentos são extraídos de um SGBD para um arquivo texto contendo as colunas (que representam os campos das tabelas) separadas por tabulações. Depois da geração do arquivo texto, dados sem significado são retirados, configurando a etapa de limpeza dos dados. Após a limpeza, os dados sofrem uma codificação que os enriquecem e os preparam para o processo de descoberta.

Para a realização dos experimentos foi utilizado um micro computador AMD K6-II com 256MB de memória RAM e 18GB de espaço em disco para analisar 63.534 registros (número total de registros disponibilizados para testes) de chamadas telefônicas do período entre 01 de setembro a 31 de dezembro de 350 assinantes inadimplentes de telefones fixos, com os seguintes dados: dia da semana que a chamada foi executada, hora inicial da chamada, destino da chamada que irá identificar o tipo do destino (local, DDD, Celular, DDI) e duração das chamadas.

Para os testes o atributo origem das chamadas não será considerado, pois não interfere na obtenção do padrão geral dos usuários em questão visto que não contribui consideravelmente como atributo delimitador de perfil geral de inadimplentes. Deve-se considerar o atributo origem no momento que especifica-se o perfil individual do inadimplente.

A utilização de um número maior de informações sobre inadimplência não foi possível

devido a este tipo de registro ser sigiloso e ponto estratégico das empresas no combate a inadimplência.

3.3 Transformação dos dados

Após a limpeza dos dados, é necessário realizar a codificação dos mesmos, ou seja, deve-se criar padrões para representar os dados. A Tabela 1 exemplifica a codificação dos dados.

Dia da semana	Hora	Destino	Duração	Degrau
1	2	1	1	1
2	1	2	2	1
3	3	2	3	2
4	5	3	4	3
5	6	3	5	4

Tabela 1: Exemplo de codificação dos dados

Os atributos foram codificados conforme segue:

- Dia da semana: esta coluna trás a data (dd/mm/aa) da realização da chamada mas foi alterada para dia da semana (segunda feira, terça feira, etc.), para que se possa saber em qual(is) dia(s) da semana os usuários tem preferência para utilização do telefone.;

Dia da semana	Codificação
Domingo	1
Segunda-feira	2
Terça-feira	3
Quarta-feira	4
Quinta-feira	5
Sexta-feira	6
Sábado	7

Tabela 2: Transformação do atributo Dia da semana

- Hora em que a chamada foi realizada. Este parâmetro foi dividido em 04 partes para especificar o período do dia que é de preferência dos usuários para a geração das chamadas. Não foi utilizada a divisão sugerida pelo sistema de tarifação das operadoras de telecomunicações, pois este especifica períodos diferenciados de cobrança para finais de semana, feriados e dias úteis;

Hora	Codificação
00:00h – 06:00h	1
06:00h – 12:00h	2
12:00h – 18:00h	3
18:00h – 24:00h	4

Tabela 3: Transformação do atributo Hora

- O atributo duração foi codificado para apresentar o parâmetro que permite saber qual a média de tempo que os usuários falam durante as conversações, foi criada uma classe para agrupar as chamadas que tem uma duração relativamente curta (menor que 60s), outra para representar chamadas com duração média (entre 61s e 120s), uma classe para definir chamadas com duração de média para alta (entre 181s e 300s), e uma classe que agrupa as chamadas com longa duração (acima de 300s);

Duração (s)	Codificação
00 – 60	1
61 – 180	2
181 – 300	3
300 – 9999	4

Tabela 4: Transformação do atributo duração

- O atributo destino foi codificado para especificar se a chamada foi realizada para um telefone fixo local, para um telefone celular local, para um telefone fixo interurbano, etc.

Destino	Codificação
Fixo (Local)	1
Celular (Local)	2
Fixo (DDD)	3
Celular (DDD)	4
Fixo (DDI)	5
Celular (DDI)	6
Serviços especiais	7

Tabela 1: Codificação do atributo destino

Estes dados foram todos codificados utilizando uma planilha e salvos em formato .txt com as colunas separadas por tabulação.

3.4 Análise utilizando árvores de decisão

Para análise utilizando árvores de decisão utiliza-se o software *Sipina* **Error! Reference source not found.** foi selecionado para a implementação das técnicas de *Data Mining*, porque possui licença para uso educacional, implementa o método de classificação e utiliza árvore de decisão para representar o conhecimento obtido. Foi desenvolvido para trabalhar com o sistema *Windows*. É um software que extrai o conhecimento a partir de dados armazenados em arquivos texto ou em formato proprietário (arquivo gerado através da própria ferramenta), e a sua capacidade de aprender está relacionada à quantidade e qualidade dos dados.

Durante a análise os resultados podem ser visualizados nas diferentes janelas do *Sipina*, como a

janela gráfica, a matriz de classificação e as regras. Pode-se citar a capacidade da ferramenta de gerar conhecimento na forma de Regras de Produção, através do aprendizado baseado em exemplos caracterizados por dados simbólicos e numéricos. Também é capaz de testar o conhecimento através de exemplos não utilizados no aprendizado e adicionar o conhecimento gerado automaticamente ao conhecimento dado por um especialista humano e otimizar a base de conhecimento **Error! Reference source not found.**

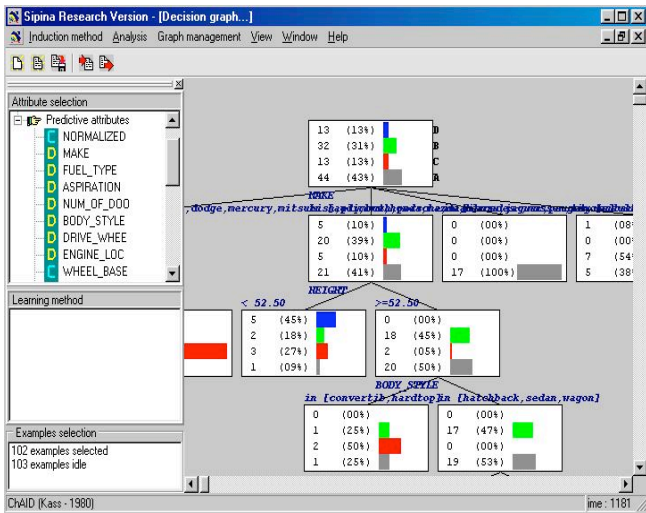


Figura 1: Interface principal do Sipina

Após a limpeza e codificação dos dados, os mesmos foram importados para o software *Sipina*, como mostra a Figura 2:

Data	Hora	Origem	Destino	DURACAO
1	3.00	6.00	1.00	1.00
2	3.00	6.00	1.00	1.00
3	3.00	5.00	1.00	1.00
4	3.00	6.00	1.00	1.00
5	3.00	6.00	1.00	1.00
6	3.00	6.00	1.00	1.00
7	3.00	5.00	1.00	1.00
8	3.00	5.00	1.00	1.00
9	3.00	5.00	1.00	2.00
10	3.00	6.00	1.00	3.00
11	3.00	6.00	1.00	2.00
12	3.00	6.00	1.00	1.00
13	3.00	5.00	1.00	1.00
14	3.00	6.00	1.00	1.00
15	3.00	9.00	2.00	1.00
16	3.00	6.00	1.00	1.00
17	3.00	9.00	2.00	1.00
18	3.00	5.00	1.00	1.00
19	3.00	7.00	1.00	1.00
20	3.00	8.00	1.00	1.00
21	3.00	6.00	1.00	1.00
22	3.00	7.00	1.00	1.00
23	3.00	5.00	1.00	1.00

Figura 2: Dados importados para o SIPINA

Nesta etapa foram adicionadas a todos os atributos as descrições de acordo com a especificação da codificação. A Figura 3 apresenta a forma de edição das categorias (atributos) utilizando o software *Sipina*.

Name	Type	Categories	Show	Decimals
1 Data	Discrete	5	Description	2
2 Hora	Continuous		Description	2
3 Origem	Continuous		Description	2
4 Destino	Continuous		Description	2
5 DURACAO	Continuous		Description	2

Description	Code
Domingo	0
Segunda-feira	1
Terça-feira	2
Quarta-feira	3
Quinta-feira	4
Sexta-feira	5
Sabado	6

Figura 3: Edição das categorias no SIPINA

Depois de efetuada a padronização dos dados no software, foram geradas algumas árvores de decisão.

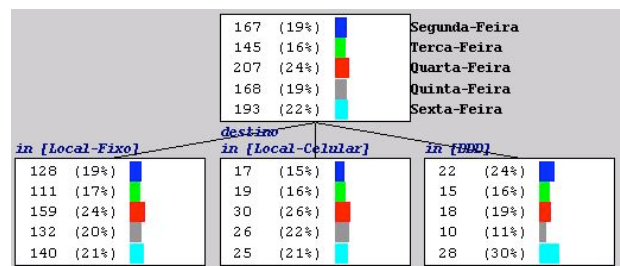


Figura 4: Exemplo de árvore de decisão gerada pelo SIPINA

3.5 Resultados obtidos com o SIPINA

Para a geração da árvore de decisão representada na Figura 4 foi montado o seguinte cenário: o algoritmo utilizado foi o C4.5 para geração da árvore de decisão, a classe principal foi criada com o atributo “dia da semana” como nó principal, com a finalidade de descobrir quais os dias da semana são preferidos dos usuários, e, os nós filhos especificados foram determinados com os atributos: hora e destino (tipo da chamada) para verificar qual o horário durante os dias da semana prefere-se gerar chamadas e qual é o tempo utilizado nestas conversações. Como resultado as seguintes regras geradas pela árvore de decisão são consideráveis.

- Regra 01: Os dias da semana com maior número de chamadas são quarta e quinta feira no horário entre 06:00h – 12:00h:
 Quarta feira: 22%
 Quinta feira: 31%
 Na sexta feira o horário de maior tráfego é entre 18:00h – 24:00h: 21%
- Regra 02: Nas segundas feiras o horário entre 12:00h e 18:00h concentra chamadas para serviços especiais: 33%
- Regra 03: Nas quartas feiras o horário entre 12:00h e 18:00h concentra chamadas:
 para telefone fixo (Local): 23%
 para telefone celular (DDD): 22%
- Regra 04: Nas quintas feiras o horário entre 12:00h e 18:00h concentram-se chamadas para telefone fixo (DDI): 30%
- Regra 05: Nas sextas feiras o horário entre 12:00h e 18:00h concentram-se chamadas:
 para telefone celular (Local): 22%
 para telefone fixo (DDD): 24%

As regras acima sugerem o perfil geral dos usuários encontrado no volume de dados avaliado. A definição do perfil geral determina o comportamento generalizado dos usuários inadimplentes. Essa definição serve de parâmetro comparativo na verificação de semelhança do perfil do usuário individual com o perfil do grupo de inadimplentes. Quando da análise dos dados de um usuário em específico, pode-se comparar o resultado com as classes pré-determinadas para verificar se este usuário encaixa-se em um dos perfis já encontrados.

Analisando os dados de um único usuário pode-se verificar se ele é pertencente a uma determinada classe já descoberta utilizando índices de aproximação.

Um usuário não pode ser considerado inadimplente / fraudador somente devido ao seu perfil enquadrar-se em uma classe pré determinada, por isso a intervenção humana (analista de fraudes) é de fundamental importância. Para evitar equívocos, os dados que geram alarmes (perfil que encaixa-se no grupo de inadimplentes) são analisados por um analista de fraudes que tem condição de realizar um estudo mais apurado sobre aquele usuário especificamente, para daí sim determinar se é ou não inadimplente / fraudador.

3.6 Análise dos dados com redes neurais

Nesta seção será descrito o experimento com redes neurais e para isso utilizou-se o software *QwikNet* que é um simulador para redes neurais que executa diversos métodos eficientes para treinar e testar uma rede neural.

A rede neural executada no *QwikNet* é chamada de *perceptron multi-layer*, que é uma rede composta de diversas camadas. O programa *QwikNet* oferece uma relação flexível e intuitiva que permite projetar, treinar e testar redes neurais em um ambiente gráfico. Criar aplicações feitas sob encomenda com o *QwikNet* é fácil. Permite projetar a rede usando a interface gráfica e então exportá-la em um de diversos formatos populares incluindo: C/C++, Pascal ou Matlab **Error! Reference source not found.**

Características do QwikNet **Error! Reference source not found.:**

Learning Rate : A taxa de aprendizado controla a taxa à qual a rede aprende. Normalmente, quanto mais alta a taxa de aprendizagem, mais rápido a rede aprende. Se a taxa de aprendizado é muito alta, a rede pode ficar instável.

Momentum: Este parâmetro controla a influência da última mudança de peso na atualização de peso atual. Normalmente resulta numa aprendizagem mais rápida, mas pode causar instabilidade em alguns casos se for muito grande.

Weights: Os campos Mínimo e Máximo permitem ao usuário fixar os mais baixos e as mais altas variâncias

para cada peso. Durante o treinamento, todos os pesos são usados conforme estes parâmetros.

Hidden layers: A rede pode conter até 5 camadas escondidas, cada uma com qualquer número de neurônios. Cada camada escondida pode consistir de qualquer função de ativação: linear, *logistic*, tangente hiperbólica, ou Gaussian. A escolha da função de ativação depende da natureza dos dados de treinamento. *Logistic* é a escolha mais comum, mas a tangente e a hiperbólica podem resultar em treinamentos mais rápidos para muitos problemas. O melhor tamanho para uma rede depende do problema.

3.6.1 Resultados obtidos com redes neurais

Para a realização dos testes utilizando redes neurais, os dados foram codificados conforme especificado no item de 3.3 Transformação dos **dados**.

A rede neural foi treinada utilizando um arquivo com 499 linhas de dados linhas (500 é o número máximo de linhas permitida pelo software *QwikNet* em caráter de demonstração) com informações (bilhetes com os atributos dia da semana, horário, destino da chamada, duração) de usuários inadimplentes, e, para o teste foi utilizado um arquivo com 240 linhas com informações de um único usuário para visualizar se este tem um perfil semelhante ao que a rede neural conseguiu aprender.

O teste foi realizado com 04 (quatro neurônios) de entrada mais um neurônio de saída, a taxa de aprendizado de 0,1, momentum 0 e critério de parada com 25.000 épocas de treinamento. A Figura 5 apresenta os resultados dos testes:

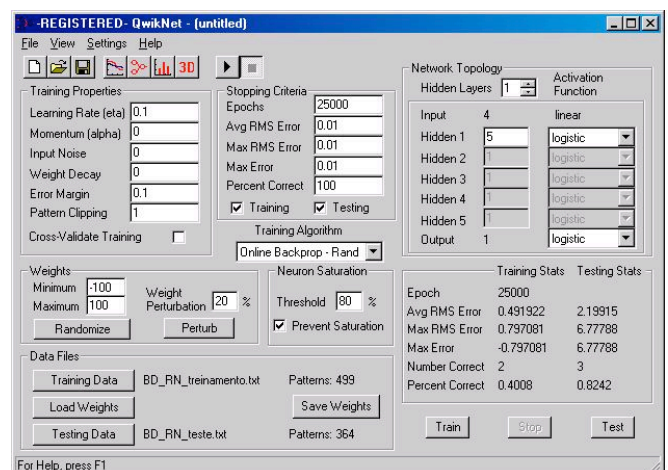


Figura 5: Interface do QwikNet

Do total de registros (364) submetidos ao teste, três registros de chamadas encaixam-se no perfil aprendido pela rede. Um indicador bastante baixo para

especificar que este é um usuário com tendências a inadimplência / fraude.

Ferramentas que utilizam redes neurais para a detecção de fraudes começam a aparecer mais no mercado. A Br Telecom investiu em 2004 um valor estimado entre US\$ 1,5 milhão e US\$ milhões em uma solução combinada da Agilent Technologies e Cerebrus para monitorar em tempo real a sua rede local, de longa distância e GSM **Error! Reference source not found.**

As redes neurais são utilizadas para aprender com o histórico dos usuários analisando o comportamento de cada um em diferentes períodos do dia. Como fazem a análise individual dão condições às empresas de descobrir em tempo real alguma atividade suspeita. A análise destes dados dá condições de interromper a atividade fraudulenta rapidamente, colaborando assim para o aumento da lucratividade da empresa **Error! Reference source not found.**

4 Comparação de resultados entre árvores de decisão e as Redes Neurais

Árvores de decisão têm uma estrutura fácil de aprender e assimilar apresentando os resultados de uma forma mais clara para um analista de resultado. As árvores de decisão apresentam os resultados em forma de árvore, facilitando a leitura. Um ponto a ser observado é que a árvore de decisão utiliza-se somente de um atributo para dividir cada nível, isso solicita a escolha correta desses atributos para que seja gerada uma árvore com um número reduzido de subconjuntos.

As redes neurais são capazes de aprender através de treinamentos específicos. As RNAs trabalham somente com números, isso obriga a transformação de determinados tipos de dados para um padrão que a RNA possa trabalhar. A obtenção do resultado muitas vezes torna-se nebulosa pois não se tem a forma exata que técnica utilizou para chegar ao resultado, além de dispensar um tempo para treinamento / aprendizagem da rede.

5 Conclusão

Os índices de inadimplência no setor de telecomunicações superam a marca de R\$ 2 bilhões [4], e em alguns casos esse valor perfaz em média 10% da receita bruta das operadoras de telecomunicações.

A inadimplência cresce mais a cada dia no setor de telecomunicações, com isso as empresas do setor deixam de arrecadar recursos de grande importância para a melhoria de seus serviços. Com a preocupação de diminuir os índices de inadimplência e melhorar a qualidade de atendimento aos seus clientes, as operadoras de telecomunicações buscam identificar o

perfil dos usuários inadimplentes e também realizar um trabalho de prevenção a fraudes.

Com a descoberta do perfil do inadimplente as operadoras de telecomunicações tem condições de fazer um trabalho preventivo de combate a falta de pagamento. Com os perfis definidos, é possível realizar uma análise sobre todas as chamadas telefônicas e verificar quando uma chamada que está em curso encaixa-se em um desses perfis, gerando alarmes para a operadora de telecomunicações e para o próprio usuário do serviço antes que este venha a ser suspenso. Também pode-se usar o conceito de definir perfil para descobrir os bons clientes dando possibilidade da realização de um trabalho de marketing direcionado à essas pessoas que hoje em dia fazem do telefone uma ferramenta de grande utilidade no seu dia a dia.

As técnicas utilizadas para mineração de dados foram: árvores de decisão e redes neurais. Ambas mostraram-se eficientes na realização do trabalho, porém, apresentaram alguns detalhes consideráveis.

A técnica de árvore de decisão apresentou dados que indicam um padrão de comportamento / perfil, porém, gerou um grande número de subdivisões tornando a leitura do resultado pouco ágil / demorada, mas de fácil compreensão. Também permite a geração das regras que definiram o padrão, isso facilita a busca por novos padrões em um novo volume de dados submetidos a esta técnica.

As RNs apresentaram um resultado um pouco opaco não dando condições de avaliar a técnica e quais os critérios utilizados para chegar ao resultado final. Desta forma, não pode-se especificar qual o critério que definiu o padrão dos usuários inadimplentes, isso é de conhecimento somente da RN.

Trabalhos futuros

A partir deste trabalho, percebe-se que um estudo mais aprofundado pode ser dirigido para:

- analisar qual a técnica que pode ser implementada com um baixo custo computacional para disponibilizar o recurso para empresas menores;
- analisar qual a contribuição que a combinação das duas técnicas apresentadas no trabalho (árvores de decisão e redes neurais) trazem para a descoberta da inadimplência;
- analisar qual a contribuição que a combinação de outras técnicas de mineração de dados trazem para a descoberta da inadimplência, e;
- estudar os benefícios da análise de inadimplente(s) / fraudadores no menor tempo possível para as empresas de telecomunicações.

As respostas a estas questões, a serem abordadas em trabalho posterior, são de fundamental importância na descoberta e combate aos inadimplentes.

Referências

- [1] AMO, Sandra. Técnicas de Mineração de dados. Anais do XXIV Congresso da Sociedade Brasileira de Computação. Integração e diferenças Regionais: O papel da Computação. Volume 2. Livro texto. 31 de julho a 06 de agosto de 2003. Salvador – Bahia.
- [2] CARVALHO, Luís Alfredo Vital de. DATA MINING. A mineração de dados no Marketing, Medicina, Economia, Engenharia e Administração. 2^a Edição. Ed. Érica. 2002
- [3] DE FRANCESCHI, AnaLúcia Schiaffino Morales. Aplicação de técnicas de inteligência artificial no desenvolvimento de agentes para gerência de redes. Tese de Doutorado. UFSC. 2003.
- [4] Disparada da inadimplência mobiliza operadoras. Revista World Telecom. Ed. 44. 2002. <http://worldtelecom.uol.com.br/AdPortalV3/adCmsDocumentoShow.aspx?Documento=27233>. Acessado em 02/08/2005
- [5] DIREITO EM DEBATE. Direitos e Deveres sobre telefonia. Artigo. <http://www.direitoemdebate.net/constelefonia.html>. Acessado em 02/08/2005
- [6] FLORES, Christian Feltrin. Ferramentas de Data Mining. 2003 <http://www.hcaa.com.br/antiga/dw/capa113123.htm>. Acessado em 13/08/2005.
- [7] GONÇALVES, Alexandre Leopoldo. Utilização de técnicas de mineração de dados em bases de dados C&T: Uma análise dos grupos de pesquisa no Brasil. Dissertação de Mestrado. UFSC. 2000
- [8] LÜDKE, Vitor Hugo. Redes Neurais Artificiais BackPropagation – QwikNet. ULBRA. 2001. <http://www.ulbra.tche.br/~danielnm/ia/trabs2000-1/vitor/trabg2.htm>. Acessado em: 28/09/2004
- [9] NERCESSIAN, Olívia. BrT Investe em solução de combate à fraude. Telecom WEB. Artigo. 2004. <http://www.telecomweb.com.br/noticias/artigo.asp?id=77560>. Acessado em: 02/12/2004
- [10] NETO, Vicente Soares; CARVALHO, Francisco Teodoro Assis. Telecomunicações. Tecnologia de centrais telefônicas CPA-E, CPA-T, Tráfego e Sinalização. 1^a Edição. Editora Érica. 1999.
- [11] OLIVEIRA, Ana Paula. Br Telecom investe em redes inteligentes para combate à fraudes. World Telecom Network World. Artigo. 2004. <http://worldtelecom.uol.com.br/adCmsDocumentoShow.aspx?GUID=5C3346FE-460C-4FC7-87EA-25FF62CF08F5&ChannelID=40>. Acessado em: 01/12/2005
- [12] OLIVEIRA, Clarissa. Empresas retomam investimento em tecnologia, diz pesquisa. Agência Estado. Artigo. 2004 <http://www.estadao.com.br/tecnologia/telecom/2004/jun/22/192.htm>. Acessado em 02/08/2004.
- [13] PINHO, Paulo. Sem desperdício de dinheiro. Revista RNT. Edição março de 2003. <http://www.rnt.com.br/edicao0305/opinio.asp>. Acessado em 31/08/2004
- [14] QWIKNET. Professional Neural Network Software [.http://qwiknet.home.comcast.net](http://qwiknet.home.comcast.net). Acessado em: 28/09/2005.
- [15] RAKOTOMALALA, Ricco. Sipina Overview. <http://eric.univ-lyon2.fr/~ricco/sipina.html>. Acessado em: 10/10/2005
- [16] ROCHA, Carlos Eduardo. . 2004. O ano da concorrência. Revista RNT. Análise. 2002 http://www.rnt.com.br/edicao0401_02/analise.asp. Em 29/07/2004
- [17] SOARES, Edileuza. PT Inovação disputa projeto da Vivo. Revista RNT. Notícia. <http://www.rnt.com.br/rnt/rntNoticiasMostrar.asp?id=615>. Acessado em 02/08/2005
- [18] SOUSA, Mauro Sérgio Ribeiro de . Mineração de Dados: Uma implementação fortemente acoplada a um sistema gerenciador de banco de dados paralelo. Tese UFRJ . Agosto de 1998
- [19] WALDMAN, Hélio; YACOUN, Michel Daoud. Telecomunicações. Princípios e Tendências. 3^a Edição. Ed. Érica. 1999
- [20] ZANUSSO, Maria Bernadete. Data Mining. http://www.dct.ufms.br/~mzanusso/Data_Mining.htm. Acessado em 23/11/2005