5CONTECSI June 04 - 06, 2008 São Paulo, Brazil

5th International Conference on Information Systems and Technology Management
5º Congresso Internacional de Gestão da Tecnologia e Sistema de Informação
De 04 a 06 de Junho de 2008 - São Paulo - Brasil

PS-1098

DECISION SUPPORT SYSTEM FOR PREVENTING *NO-SHOW* TO MEDICAL APPOINTMENTS

Paulo J.L. Adeodato (NeuroTech and Universidade Federal de Pernambuco, PE, Brazil) – paulo@neurotech.com.br and pjla@cin.ufpe.br
Adrian L. Arnaud (NeuroTech and Universidade Federal de Pernambuco, PE, Brazil) – adrian@neurotech.com.br and ala2@cin.ufpe.br
Victor M. Braz (NeuroTech, PE, Brazil) – victor@neurotech.com.br
Germano C. Vasconcelos (NeuroTech and Universidade Federal de Pernambuco, PE, Brazil) – germano@neurotech.com.br and gcv@cin.ufpe.br

The failure of customers to attend booked appointments, known as *no-show*, is a problem faced by companies worldwide, operating in medical assistance and transportation businesses. This work focused on the former, more particularly, on the medical appointments made via call-center. This paper presents a decision support system based on data mining for identifying, at booking time, the medical appointments with high risk of *no-show*, for helping online re-scheduling. Pre-processing and data transformation yielded embedding expert's knowledge and behavioral information. The a priori algorithm explicited the knowledge contained on the data and an MLP neural network estimated the risk of *no-show*. The system has been developed on a data set of 30,000 and tested on other 10,000 appointments from a healthcare company operating in Brazil. Both the risk estimation and the rules extracted attained high quality in the metrics defined and were considered very relevant by the company's specialist.

Keywords. No-show in medical assistance, Data mining, Decision support systems, Neural networks, A priori algorithm.

PREVISÃO DE *NO-SHOW* NO AGENDAMENTO DE SERVIÇOS MÉDICOS BASEADA EM MINERAÇÃO DE DADOS

O não comparecimento de clientes a compromissos agendados, o *no-show*, é um problema típico para as empresas de assistência médica e de transporte de passageiros. Este trabalho focou no agendamento de consultas médicas feito via *call-center*. Este artigo apresenta um sistema de apoio à decisão baseado em mineração de dados para identificar as consultas com alto risco de *no-show* para auxiliar seu re-agendamento em tempo real. O objetivo foi estimar e minimizar a perda com a alocação de recursos causada pelo *no-show*. O conhecimento do especialista e as variáveis de comportamento foram incorporados à visão dos dados, uma rede neural MLP estimou o risco de *no-show* e o algoritmo a priori explicitou as regras sobre o negócio. O sistema usou 30.000 consultas para modelagem e outras 10.000 para teste, da base de uma empresa de assistência médica. O desempenho da rede neural e a qualidade das regras atingiram um alto nível nas métricas usadas e foram de grande importância para o especialista.

Palavras-Chave: *No-show* em assistência médica, Mineração de dados, Sistemas de apoio à decisão, Redes neurais, Algoritmo a priori.

1. Introdução

O problema do não comparecimento a compromissos agendados implica em grandes custos às empresas prestadoras de serviços pela alocação desnecessária dos seus recursos. Este problema, comumente chamado de *no-show*, ocorre freqüentemente em serviços de assistência médica e de transporte de passageiros, sem qualquer comunicação prévia da impossibilidade de comparecimento por parte do cliente.

Neste cenário, o provedor do serviço tem que dimensionar a alocação dos seus recursos de forma a minimizar as perdas balanceando aquelas causadas por disponibilizar recursos não utilizados com aquelas causadas pela falta de recursos para atender aos clientes.

Este trabalho apresenta um estudo realizado sobre o problema de *no-show* em consultas médicas a partir do agendamento via call-center, para uma empresa prestadora de serviços médicos numa grande cidade brasileira. Foram utilizadas técnicas das mais avançadas cientificamente para desenvolver um sistema de apoio à decisão na estimação do risco de *no-show* com base em uma amostra das consultas agendadas pelos pacientes num determinado período, como descrito a seguir.

O artigo está organizado em 7 seções da seguinte forma. A Seção 2 faz uma revisão da literatura médica sobre o tema. A seção 3 caracteriza o problema em termos de objetivos, respostas desejadas, variáveis de entrada e massa de dados disponível. A Seção 4 caracteriza a massa de dados e apresenta o tratamento dos dados até a codificação das variáveis. A Seção 5 descreve a indução das regras que representam o conhecimento que pode ser explicitado embutido nos dados e o desempenho obtido com essa técnica. A Seção 6 mostra a estimação do *no-show* por meio de uma rede neural e o seu desempenho. A Seção 7 tece comentários finais sobre o trabalho desenvolvido, as suas limitações e o que pode ser realizado em termos de aprimoramento e de serviços adicionais.

2. Revisão da literatura

O problema de *no-show* no agendamento de consultas tem afetado os serviços médicos de diversos países e tem sido estudado sob variados pontos de vista. Muitos trabalhos têm analisado as causas do problema [Moore *et. al*. 2001], [Stone *et. al*. 1999]; outros têm avaliado as conseqüências econômicas [Bigby *et. al*. 1983] ou médicas [Bigby *et. al*. 1984] do problema. Alguns deles buscam potenciais ações para resolver o problema e analisam a viabilidade econômica dessas ações [Bigby *et. al*. 1983] .

Os níveis de *no-show* apresentam uma grande variação (em geral, de 20% a 60% das consultas agendadas) por razões de diversas naturezas que vão desde as origens culturais do paciente até a falta de transporte público adequado, passando pelo esquecimento do próprio paciente. Vários desses estudos julgam que estas causas variam fortemente de acordo com o grupo de pacientes estudado.

Um estudo sobre o agendamento via call-center das consultas no Medical Center de Augusta, VA, USA [Maclean *et. al*. 2001] mostrou que o uso de enfermeiras como operadoras do call-center em 1999, reduziu a taxa de *no-show* de 21% do método convencional para 5,5%. Considerando o custo de US$ 65.13 por *no-show*, o estudo estimou em US$ 328,134.00 a redução das perdas do hospital com *no-show*, caso todos os agendamentos daquele ano tivessem sido feitos dessa forma.

Estudo similar sobre a viabilidade econômica em se fazer lembrete aos pacientes [Bigby *et. al*. 1983], mostrou que o lembrete é efetivo na redução do *no-show*, tanto via telefonema quanto via carta. O estudo da relação custo/benefício mostrou, porém que a ação só se torna economicamente viável para aquele serviço quanto a taxa de *no-show* esperada estiver acima de 5% para lembretes via carta ou 11% para lembretes telefônicos.

A grande maioria dos estudos realizados [Bigby *et. al*. 1983], [al-Khadra *et. al*. 1992], [Gruzd *et. al. 1986*] abordou o problema de *no-show* em domínios muito restritos e sobre amostras reduzidas de pacientes, em estudos controlados, de acordo com os paradigmas habitualmente utilizados na Medicina. Poucos desses estudos, porém utilizam técnicas estatísticas de análise multivariada [Gruzd *et. al*. 1986]. A abordagem aqui proposta, por outro lado, é aquela da mineração de dados, onde, as grandes massas de dados registradas durante o funcionamento rotineiro dos serviços médicos contêm muita informação útil a ser extraída pelas técnicas multivariadas de Inteligência Artificial e Estatística para apoiar a tomada de decisão.

3. Caracterização do problema

O agendamento de um serviço se inicia com a ligação do paciente para o call-center. A operadora digita a identificação do paciente e sua solicitação de agendamento com especificação de data e horário, especialidade ou médico e unidade hospitalar. A operadora verifica a elegibilidade do paciente ao serviço e lhe comunica qualquer impossibilidade de atendimento àquela especificação de demanda, oferecendo horários ou médicos alternativos ou qualquer outra alteração que se faça necessária para atender ao paciente.

É importante frisar que o agendamento é direcionado pelo desejo do paciente, mantidas as restrições de disponibilidade do serviço médico solicitado. Essa liberdade para o paciente, entretanto, em diversas vezes, o leva a um agendar a consulta em condições inconvenientes para o seu comparecimento. Um dos objetivos deste trabalho é identificar que condições são essas que freqüentemente causam a situação de *no-show*.

Na caracterização do problema, a empresa estabeleceu que o *no-show* ocorre quando o paciente não comparece à consulta ou se atrasa em mais de uma hora em relação ao horário marcado. Essa caracterização transforma a previsão de *no-show* em um problema de classificação binária. O horário de comparecimento do paciente que é registrado na base de dados pelas atendentes do serviço médico serve para caracterizar ocorrência de *no-show* por excesso de atraso.

Na realidade, uma consulta agendada resultava em três possíveis situações: comparecimento, falta ou cancelamento antecipado. Como o cancelamento da consulta não causa prejuízo ao serviço médico, ele foi excluído das consultas. Entretanto, o cancelamento manifesta o comportamento do cliente e do atendente em relação ao *no-show*, como será visto adiante. Por isso, as consultas canceladas foram computadas nas variáveis comportamentais das demais consultas, no processo de modelagem, antes de serem descartadas.

A alocação dos recursos não utilizados por causa do *no-show* produz uma perda cujo valor não foi especificado nem relacionado pela empresa ao valor da perda de clientes causada pelo *over-booking* do serviço. O conhecimento do impacto em perdas serviria como importante critério de otimização dos parâmetros de controle da operação do sistema.

4. Massa de dados

A massa de dados consistia de cerca de 46.500 consultas agendadas pelo call-center da empresa de assistência médica numa grande cidade brasileira. Das três possíveis situações das consultas agendadas, cerca de 6 mil consultas eram canceladas, 12,5 mil eram efetivadas e as 28 mil restantes resultaram em *no-show*.

A base continha 14 variáveis explicativas (de entrada) e a variável alvo. Algumas variáveis e formas de preenchimento foram eliminadas ou codificadas por questões de confidencialidade, restando 12 variáveis. Tais variáveis eram diretamente extraídas da base de dados relacional enquanto outras eram variáveis calculadas para inserção de conhecimento do especialista ou para captura do comportamento do paciente e do atendente do call-center.

Foram criadas 3 variáveis a partir das datas cadastradas no banco de dados. O *Tempo_plano* é o valor em dias entre a data de admissão no plano e a data de marcação da consulta, o *Tempo_Consulta* é a diferença em dias da data consulta para a data de marcação e a variável *Idade* representa a idade do usuário na marcação da consulta. A variável *Status*, que possuía a informação *a posteriori,* foi categorizada em uma variável dicotômica (*No-show*) indicando se houve ou não o comparecimento do paciente à consulta agendada pelos critérios descritos na seção anterior. Essa categorização resultou no alvo da função-objetivo da aprendizagem supervisionada, tanto para a estimação do risco de *no-show*, quanto para a indução das regras de classificação. Cerca de 50 variáveis adicionais foram criadas para capturar o comportamento (behavior) do paciente e da atendente em relação ao histórico até a data de agendamento de cada consulta. No horizonte histórico, o comportamento era consolidado em 15 dias, 1 mês e toda a história do paciente ou da atendente.

As 18 variáveis de maior importância usadas na solução do problema estão na Tabela1. Essa importância é calculada a partir dos parâmetros da rede neural treinada e normalizada pela de maior valor, como será visto mais adiante.

Tabela 1. Lista de Variáveis e suas Influências na Rede Neural

Rank	Descrição	Tipo	Importância
1	QTDE_CONSULTAS_EFETIVADAS_15DIAS_ATENDENTE	behavior	1,00
2	QTDE_CONSULTAS_CANCELADAS_15DIAS_ATENDENTE	behavior	0,82
3	UNIDADE	Original	0,76
4	QTDE_CONSULTAS_CANCELADAS_1MES_ATENDENTE	behavior	0,71
5	QTDE_CONSULTAS_EFETIVADAS_1MES_ATENDENTE	behavior	0,70
6	QTDE_CONSULTAS_CANCELADAS_TOTAL_ATENDENTE	behavior	0,64
7	QTDE_CONSULTAS_ABERTAS_15DIAS_ATENDENTE	behavior	0,54
8	QTDE_CONSULTAS_EFETIVADAS_TOTAL_ATENDENTE	behavior	0,48
9	QTDE_CONSULTAS_ABERTAS_1MES_ATENDENTE	behavior	0,48
10	QTDE_CONSULTAS_ABERTAS_TOTAL_ATENDENTE	behavior	0,45
11	TEMPO_CONSULTA_MARCACAO	Transformada	0,43
12	ATENDENTE	Original	0,30
13	QTDE_SOLICITACOES_15DIAS_ATENDENTE	behavior	0,20
14	QTDE_CONSULTAS_ABERTAS_1MES	behavior	0,16
15	QTDE_CONSULTAS_ABERTAS_TOTAL	behavior	0,16
16	QTDE_SOLICITACOES_1MES_ATENDENTE	behavior	0,14
17	QTDE_CONSULTAS_ABERTAS_15DIAS	behavior	0,11
18	QTDE_CONSULTAS_EFETIVADAS_TOTAL	behavior	0,10

Em seguida, dois processos de natureza e objetivos diferentes foram realizados: a codificação dos campos categóricos em numéricos para compatibilizar com a entrada das redes neurais e a discretização das variáveis numéricas em faixas para compatibilizar com o indutor de regras. Dessa etapa resultaram 2 (duas) visões da massa de dados, uma para aplicação na Rede Neural e outra no algoritmo *A Priori*.

5. Algoritmo *a priori* e indução de regras

Das diversas técnicas de indução de regras utilizamos uma variação do algoritmo *A Priori* [Witten and Frank 2005] para indução, conjugada com a técnica de refinamento de Major e Mangano [Major and Mangano 1993], que é realizada em 3 etapas.

Na primeira etapa foi aplicado o algoritmo *A Priori* que consiste em selecionar os subconjuntos mais freqüentes (*frequent itemsets*), a partir dos de mais baixa ordem. Assim, foram induzidas as regras que satisfaziam a uma cobertura mínima e que tinham uma complexidade limitada. Adotamos uma cobertura mínima de 1% da massa de modelagem porque tanto era estatisticamente representativa quanto restringia o custo computacional a valores aceitáveis. Também, limitamos a complexidade funcional a 2 variáveis de entrada para preservar um bom padrão de interpretabilidade das regras.

A segunda etapa consistiu em, automaticamente, selecionar as regras que atendiam a níveis de cobertura mais amplos que os induzidos originalmente e a níveis de confiança de interesse (*interestingness*) para cada classe. O *lift* da regra é uma importante forma de medição automática da qualidade das regras: ele mede o aumento da confiança da regra em relação à confiança média da amostra analisada. Neste caso, optamos pelos parâmetros de 1% de cobertura e confiança acima de 60% para a classe de comparecimento (*lift* mínimo de 2,0 =60%/30%) e confiança acima de 84% para a classe de *no-show* (*lift* mínimo de 1,2 =84%/70%).

A terceira (última etapa) consiste em selecionar, semi-automaticamente, as regras que apresentam mais interesse sob os pontos de vista do especialista e da eliminação de redundância nas regras. Ao especialista, além da qualidade técnica, interessam aspectos de custo, velocidade, facilidade de obtenção das informações nas regras induzidas e a utilidade da regra no apoio ao trabalho de agendamento das consultas. As métricas de similaridade entre grandezas não determinísticas, podem auxiliar o especialista a ampliar a cobertura de algumas regras pela redução da sua complexidade funcional (e.g. passar de 2 variáveis para 1 variável) mesmo com a pequena redução no *lift* da regra originalmente induzida.

Da massa de dados analisada, apresentamos abaixo 6 (seis) exemplos regras, 3 de cada categoria, que explicitam conhecimento extraído de grande interesse sobre o *no-show* dos pacientes:

Tabela 2. Regras extraídas com o algoritmo *A Priori*.

Percentual de no-show da massa de modelagem: 70%

Regras de No-show (Ruins)				**Regras de Comparecimentos (Boas)**			
Condição	**Seleção**	**Confiança**	***Lift***	**Condição**	**Seleção**	**Confiança**	***Lift***
Atendente = Atendente #027	1,7%	99,9%	1,43	Atendente = Atendente #044	2,4%	12,2%	2,93
Tempo_Consulta_Marcacao = 25 a 30 dias	8,8%	92,1%	1,32	Qtde_Consultas_Efetivadas_1Mes_Atendente = maior que 1000	1,0%	21,9%	2,60
Qtde_Consultas_Canceladas_1Mes_Atendente = 40 a 80	9,8%	90,5%	1,29	Especialidade = Fonoaudiologia	1,2%	29,6%	2,35

Vemos que as variáveis mais relevantes estão relacionadas com as *Atendentes* do call-center, sejam aquelas diretamente ligadas a atendentes específicas ou indiretamente, pela experiência ou taxa de sucesso das mesmas. Este resultado condiz com a influência das variáveis capturada pela rede neural e com os estudos realizados no Medical Center de Augusta, VA, USA [Maclean *et. al*. 2001] que mostrou o sucesso do uso de enfermeiras como operadoras do call-center na redução do no-show. Neste trabalho, a falta de treinamento específico e a alta rotatividade das operadoras do call-center eram responsáveis pela alta variação do seu desempenho em relação ao *no-show*.

A variável *Tempo_Consulta_Marcacao* para prazos longos também é um fator de risco de *no-show*, também concordando com a influência detectada pela rede neural. Na Arábia Saudita, também se verificou que o fato de o paciente apresentar problema cardíaco é estatisticamente significativo para o aumento do *no-show* [al-Khadra *et. al*. 1992], principalmente se o *Tempo_Consulta_Marcacao* for longo. Nesses casos é comum um mal-estar cardíaco levar o paciente à emergência antes da consulta agendada que, posteriormente, não é desmarcada por esquecimento de cancelamento.

Já outras *especialidades*, como fonoaudiologia, fisioterapia e outras que exigem freqüência e assiduidade para o sucesso do tratamento médico apresentam nível de comparecimento bem acima da média.

6. Redes neurais e estimação do *no-show*

Para fazer a avaliação do nível de *no-show* nas consultas agendadas, optamos pelo modelo de rede neural M*ultilayer perceptron* (MLP) treinada com o algoritmo de *backpropagation* [Haykin 1998], [Rumelhart *et al.* 1986], o qual tem sido bastante utilizado com sucesso na solução de problemas de classificação de padrões. Dentre as características mais atrativas deste tipo de rede neural é possível destacar a sua excelente capacidade de generalização, a simplicidade de operação da rede e o fato de a mesma ser um aproximador universal de funções [Hornik *et al.* 1989] e produzir uma resposta contínua que permite uma decisão baseada em limiar.

6.1. Conceitos fundamentais de redes neurais

A rede MLP é formada por conjuntos de unidades de processamento (usualmente chamados de neurônios artificiais) agrupadas em camadas. Em uma rede MLP típica estas camadas são dispostas em seqüência, com os neurônios de uma camada completamente conectados aos neurônios da camada antecessora. A cada conexão entre os neurônios é associado um peso $\omega \in \Re$. Os estímulos que passam por cada conexão (unidirecional) são ponderados por estes pesos determinando, assim, o estímulo total que chega a cada neurônio aos quais eles estão conectados. Este estímulo resultante é o argumento da função de ativação de cada neurônio artificial que gera a sua resposta àqueles estímulos. Assim, a partir de um vetor estímulo na sua camada de entrada, a rede neural transforma e propaga esses estímulos pelas suas camadas até produzir uma resposta na sua camada de saída. A função de ativação mais comumente usada em redes MLP para classificação é a função sigmóide, dada por: $y = 1 / (1+\exp(x))$. O clássico algoritmo de treinamento da rede MLP é o *error-backpropagation;* um processo iterativo de ajuste dos pesos [Haykin 1998].

Considerando que o problema foi modelado como categórico e em 2 classes, a resposta de cada neurônio pode ser conjugada por meio de uma transformação linear em uma única grandeza escalar (o escore) que definimos no domínio contínuo de 0 a 100, com o zero representando o *no-show* e o cem significando o comparecimento à consulta

marcada. Tal representação da resposta por uma grandeza escalar possibilita um monitoramento muito mais refinado do desempenho do sistema de apoio à decisão.

6.2. Treinamento da Rede Neural

As redes neurais utilizadas nos experimentos deste trabalho foram redes MLP com uma única camada escondida com 3 neurônios. Todos os pesos foram inicializados com valores aleatórios próximos a zero (da ordem de 10^{-4}) e todos os neurônios possuíam função de ativação sigmóide. Todas as redes neurais foram treinadas com o algoritmo padrão de retropropagação de erro com taxa de aprendizagem igual a 0,01 e momentum 0,1. O critério adotado para parada do treinamento foi o monitoramento do poder de generalização da rede neural utilizando um conjunto de validação (*hold out*) [Prechelt 1994] com MSE (*Mean Squared Error*) como métrica. O erro da camada de saída da rede é dado por $E_p = ½ \Sigma_{j=1} (t_{pj} - o_{pj})^2$, onde o_{pj} é a resposta do neurônio n_j e t_{pj} é a saída desejada. Diversas combinações de taxas de aprendizagem, *momentum* e neurônios da camada intermediária foram experimentadas, antes da adoção desses valores.

Todos os dados foram normalizados ou codificados de maneira que assumissem valores entre 0 e 1. A massa de dados foi separada em 3 lotes de 50%, 25% e 25%, respectivamente para treinamento, validação e teste. Para o lote de treinamento houve uma replicação dos registros de maneira a igualar a quantidade de *no-show* e comparecimento. Os lotes de teste e validação foram mantidos com as suas características originais.

A codificação das variáveis de entrada em códigos numéricos expandiu o vetor de entrada para 121 dimensões, deixando a rede MLP com 121 neurônios de entrada e 2 de saída. Na camada intermediária, foram definidos 3 neurônios, de modo a oferecer à rede a maior funcionalidade que preservasse o número de parâmetros [(2+121)x3=369] inferior ao máximo recomendado pela regra prática usual [Haykin 1998], para a quantidade de dados disponíveis no conjunto de treinamento original (cerca de 20.250),

Após o treinamento da rede neural, a influência de cada variável foi calculada e os seus valores estão indicados normalizados na Tabela 1, para as 18 variáveis mais relevantes.

5.3. Avaliação de Desempenho

A transformação da saída da rede multilayer perceptron em uma resposta contínua para o problema binário da decisão (escore) possibilita a variação do limiar de decisão (ponto de corte) para definir a separação das classes, no sistema de apoio à decisão. A escolha deste limiar depende de diversos fatores práticos reais tais como o custo relativo entre os dois tipos de erro (Tipo I e Tipo II) e o impacto da decisão.

Uma das métricas aqui usadas para medição do desempenho é o teste estatístico de Kolmogorov-Smirnov (KS2) [Conover 1999]. Este é um teste não-paramétrico usado para medir a aderência entre duas distribuições de dados a partir das suas funções de distribuição acumuladas (FDAs). Em sistemas de decisão binária baseada em limiar sobre escalar, esta métrica é usada para medir a dissimilaridade entre as distribuições. A curva de Kolmogorov-Smirnov é a diferença entre as FDAs dos dois conjuntos de dados e, quanto mais alta a curva, melhor é a sua separabilidade. O valor máximo do KS2 ($KS2_{Máx}$) é particularmente importante porque permite uma comparação direta entre sistemas decisórios distintos, nos seus pontos de melhor desempenho.

Outra das métricas aqui usadas para medição do desempenho é a curva ROC (Receiver Operating Characteristic Curve), uma ferramenta gráfica que representa o compromisso entre os verdadeiros positivos e falsos positivos resultantes de cada classificação produzida ao longo de todos os possíveis pontos de operação (limiares de decisão) do sistema decisório. Quanto mais próxima do ângulo superior esquerdo do gráfico (ponto ideal), melhor é o sistema decisório. Essa mínima distância é um importante indicador pontual do desempenho do sistema decisório; quanto menor, melhor. A curva ROC também oferece um indicador mais geral que é a área sob a curva (area under the ROC curve – AUC_ROC [Fawcett, 2003]); quanto mais próxima de um (1), melhor.

5.4. Resultados

Todos os resultados aqui apresentados foram realizados sobre o conjunto de teste com 10.196 ocorrências com independência estatística em relação aos dados de modelagem (treinamento e validação). Além disso, o conjunto de teste correspondeu a um momento futuro em relação ao período de captura dos dados para modelagem, num cenário muito realista.

Foram traçados o gráfico do KS2 e o gráfico com a distribuição das classes. A Figura 1 mostra que existe uma grande separação entre as distribuições dos escores do *no-show* e do comparecimento o que indica que o sistema é capaz de prever o *no-show* com uma boa margem de acerto. A medição do $KS2_{Máx}$ para o conjunto de teste ficou em **0,716**.

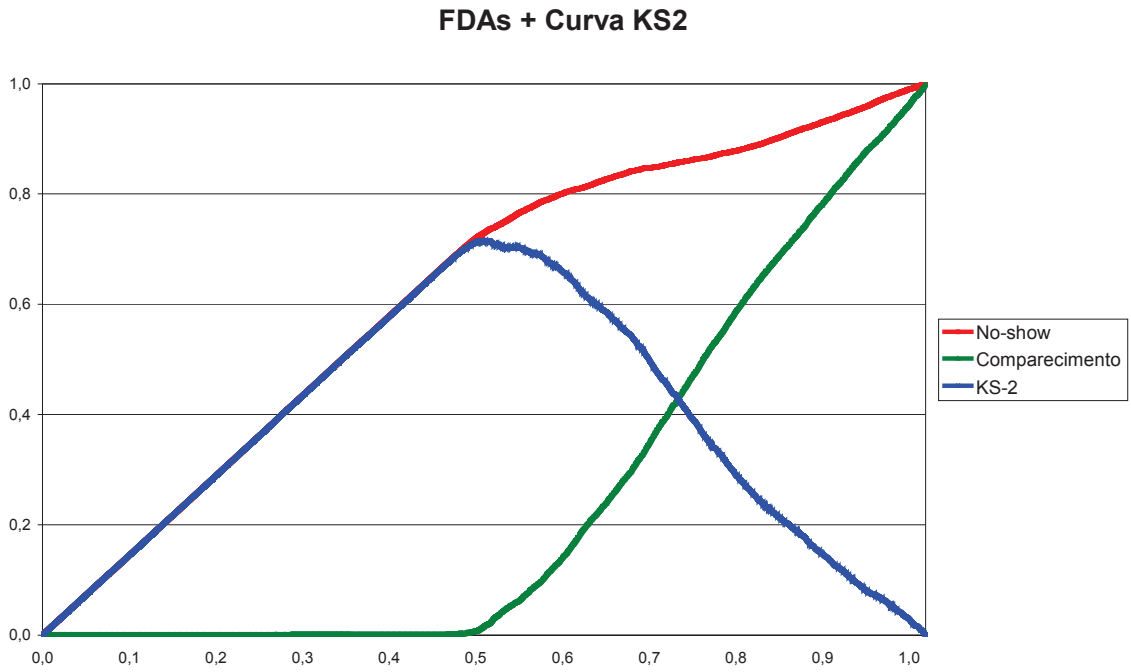

Figura 1. Funções Distribuição Acumuladas (FDAs) e Curva Kolmogorov-Smirnov (KS2).

A outra avaliação de desempenho foi realizada por meio da curva ROC. A Figura 2 mostra que existe uma grande detecção de verdadeiros positivos antes de se começar a errar nas decisões. A medida da AUC_ROC foi de **0,871** no sucesso da detecção do *no-show*.

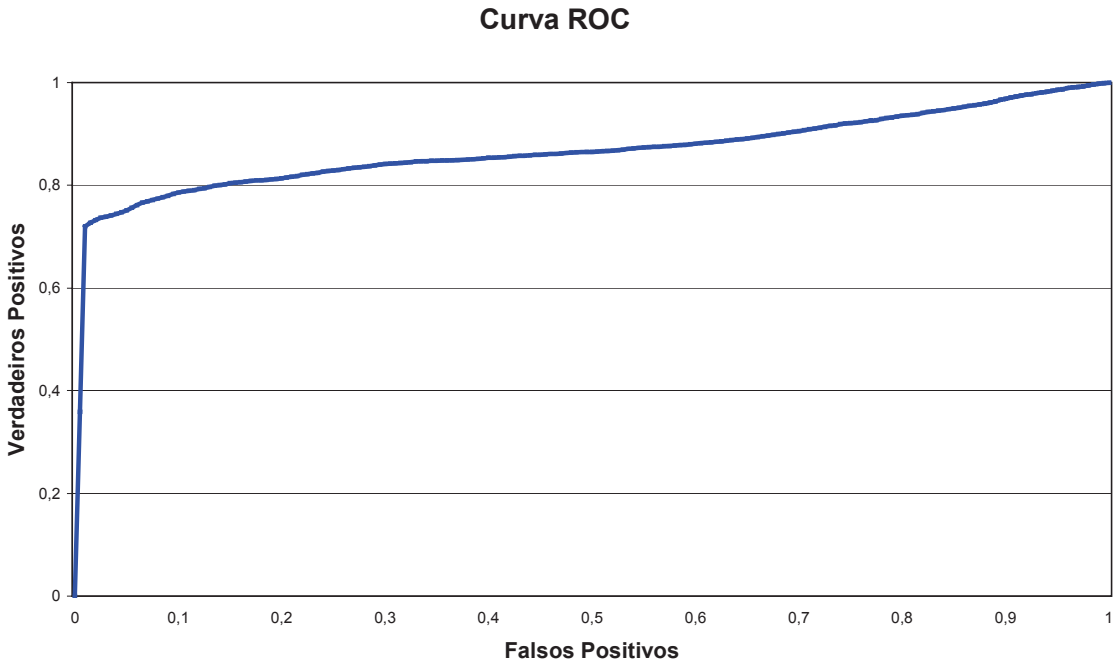

Figura 2. Curva ROC (Receiver Operating Characteristic Curve).

Apenas para exemplificar, podemos adotar o ponto de corte de máximo KS2 para operação do sistema. A fixação desse limiar produz uma decisão binária e conseqüentes erros de decisão. A Tabela-3 abaixo ilustra a matriz de confusão que mostra a taxa de acerto de 82,3% na diagonal principal e a taxa de erro de 17,7% na diagonal secundária.

Tabela 3. Matriz de Confusão no ponto de máximo KS2.

		Estimado	
		No-show	Comparecimento
Real	No-Show	53,3%	14,5%
	Comparecimento	3,2%	29,0%

6. Considerações finais

Este trabalho apresentou uma aplicação real de mineração de dados ao problema *no-show* às consultas médicas agendadas via call-center, com resultados estatisticamente relevantes quanto ao desempenho na estimação do nível de comparecimento dos pacientes. Além disso, o conhecimento sobre o problema explicitado por meio de regras induzidas automaticamente a partir dos dados foi considerado relevante por experiente especialista da empresa de assistência médica. A qualidade das regras extraídas permitiu um melhor entendimento das principais causas que levam ao *no-show* dos pacientes, possibilitando, assim, ações de re-agendamento on-line das consultas de alto risco.

A avaliação de desempenho do sistema foi realizada em uma amostra estatisticamente independente, como recomendado, porém acrescida do realismo de ser composta por dados coletados em momento cronologicamente posterior ao dos dados da

modelagem. Esse cenário típico das aplicações práticas impõe adversidades ao modelo bem mais severas que a simples amostra aleatória da mesma distribuição como comumente visto na literatura científica.

Além do perfil da solicitação de agendamento, este sistema fez análises bem mais aprimoradas em termos de perfil de paciente cadastrado e de atendente a partir dos seus comportamentos individuais (behavioral data) em consultas anteriormente agendadas.

Este sistema decisório reduziria muito *no-show*, a uma taxa abaixo mesmo daquela reportada na literatura internacional sobre o problema. Assim, o re-agendamento dessas consultas é fundamental para um apropriado dimensionamento da infra-estrutura de assistência médica. Quando da implantação do sistema em tempo real, as operadoras do call-center digitariam os dados do agendamento da consulta e veriam imediatamente o risco de *no-show* do paciente com as sugestões de alteração para tentar convence-lo a fazer um re-agendamento para redução desse risco.

Apesar de este trabalho ter sido todo realizado com o acompanhamento de um especialista da empresa de assistência médica, algumas questões, que poderiam ser mais bem tratadas e exploradas, deixaram de ser realizadas em função da dificuldade de acesso ou inexistência de informação sistematizada.

O problema investigado neste trabalho é apenas um dos muitos de interesse para o setor médico. Além disso, o problema de *no-show* tratado neste artigo, é de grande interesse para as companhias de transporte de passageiros, principalmente aéreo, onde as empresas têm que operar no ponto ótimo entre o prejuízo do *no-show* e as elevadas despesas causada pelo *over-booking* impostas pela regulamentação do setor [DAC 2003].

Referências

al-Khadra, A., Magbool, G., Wosornu, L., al-Awdah, S., Qutub, H. and Al-Khatib, R. (1992) "Why do cardiology out-patient appointments fail in Saudi Arabia?", Qual Assur Health Care, 4(4):305-10.

Bigby, J., Giblin, J., Pappius, E. M., and Goldman, L. (1983) "Appointment reminders to reduce no-show rates. A stratified analysis of their cost-effectiveness" JAMA, 7;250(13):1742-5.

Bigby, J. A., Pappius, E., Cook, E. F., and Goldman, L. (1984) "Medical consequences of missed appointments", Arch Intern Med, 144(6):1163-6.

Conover, W. J. (1999) "Practical Nonparametric Statistics", Chap. 6, Third Edition, John Wiley & Sons, New York.

DAC (2003) "Regulamentação do Transporte Aéreo de Passageiros", Divisão de Aviação Comercial — DAC, Brasil.

Fawcett, T. (2004). [17] T. Fawcett. "ROC Graphs: Notes and Practical Considerations for Researchers", 2004. Retrieved on March 2008, from http://www.binf.gmu.edu/mmasso/ROC101.pdf.

Gruzd, D. C., Shear, C. L., and Rodney, W. M. (1986) "Determinants of no-show appointment behavior: the utility of multivariate analysis", Fam Med., 18(4):217-20.

Han, J., & Kamber, M. (2006). Data Mining: Concepts and techniques. Morgan Kaufmann, San Francisco, CA.

Haykin, S. (1998) “Neural Networks: a comprehensive foundation”, 2nd ed., Prentice Hall Press PTR.

Hornik, K., Stinchcombe, M. and White, H. (1989) "Multilayer feedforward networks are universal approximators", Neural Network, Vol. 2 (5), p. 359-366.

Maclean, J. R., Armour, B., Cook, R., Etchason, J., Holmes A. K., Waller, J. L. and Clements, E., (2001) “Telephone triumph: Using a nurse-led call center to improve the coordination of medical care”, Telemedicine Journal and e-Health Summer 7(2): 184, American Telemedicine Association Sixth Annual Meeting, June 3-6, Florida.

Major, J. and Mangano, J. (1993) “Selecting among Rules Induced from a Hurricane Database”, In Proc of 1993 KDD Workshop, pp. 28--47, Washington, DC, July.

Moore, C. G., Wilson, Witherspoon P. and Probst, J. C. (2001) “Time and money: effects of no-shows at a family practice residency clinic” Family Medicine, 33(7):522-7.

Prechelt, L. (1994) “Proben1 – A Set of Neural Network Benchmark Problems and Benchmarking Rules”, Technical Report 21/94, Fakultät für Informatik, Universität Karlsruhe, Germany.

Rumelhart,D. E., Hinton, G. E. and Williams, R. J. (1986) "Learning internal representations by error propagation", MIT Press Computational Models of Cognition and Perception Series, Vol. 1, p. 318-362.

Stone, C. A., Palmer, J. H., Saxby, P. J. & Devaraj, V. S. (1999). “Reducing non-attendance at outpatient clinics”, Journal of the Royal Society of Medicine, 92(3), 117-118.

Witten, I. H., and Frank, E. (2005) “Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations”, Morgan Kaufmann, California, USA.