

Uma Aplicação de Mineração de Dados à Ativação de Usuários dos Serviços Sociais do SESC

Roberto A.F. Santos¹, Paulo J.L. Adeodato^{1,2}, Bruno P. Amorim², Carlos F. Santos²

¹ *Grupo de Inteligência Computacional, Universidade Federal de Pernambuco (UFPE)
Recife, Pernambuco, Brasil
Email: {rafs, pjla}@cin.ufpe.br*

² *NeuroTech Ltda
Recife, Pernambuco, Brasil
Email: {paulo, bruno, carlos}@neurotech.com.br*

Abstract

The Brazilian Legislation on Workers Associations defines a contribution for social services providing by Serviço Social do Comércio (SESC). Despite being free for the eligible workers and their families, these social services are not used by most of the registered users. This paper presents a decision support system based on data mining for identifying, at the moment of their enrollment, the applicants who will not use these services, allowing SESC to take pro-active measures to convince the applicants of the usefulness of these services. This system has been developed and tested on a data set of 45,000 users enrolled in Pernambuco until 2003. A multilayer perceptron neural network estimates the risk of an applicant not becoming an active service user and a rule inductor based on the A Priori algorithm explicits the knowledge contained on the data.

Keywords: Data mining, Neural networks, Decision support systems, Client activation, Database marketing.

Resumo

Este trabalho apresenta uma aplicação real de mineração de dados à ativação de usuários do Serviço Social do Comércio (SESC). O objetivo desta aplicação é possibilitar ao SESC identificar, no momento do cadastro, os usuários que não farão uso dos serviços e, como resultado, tomar uma atitude pró-ativa para convencer os usuários a utilizar os serviços. A base de dados utilizada no estudo de caso contém cerca de 45.000 usuários cadastrados nos centros de atividades

do SESC de Pernambuco até 2003. Uma rede neural foi utilizada na estimação do risco de não ativação dos novos usuários e um indutor baseado no algoritmo A Priori foi utilizado na indução de regras que propiciam um melhor entendimento do problema. Tanto o desempenho da rede neural quanto a qualidade das regras induzidas atingiram um alto nível nas métricas de avaliação utilizadas e foram considerados de grande importância pelos especialistas do SESC.

1. Introdução

O Serviço Social do Comércio (SESC) é uma organização do terceiro setor (paraestatal) que tem o objetivo de estudar, planejar e executar medidas que contribuam para o bem-estar social e a melhoria do padrão de vida do trabalhador do comércio de bens e serviços e sua família [1].

O SESC oferece aos seus associados serviços nas áreas de educação, cultura, lazer e assistência médica. Os recursos para a realização desses serviços são provenientes das contribuições parafiscais e dotações orçamentárias do poder público [2].

Em 2003, cerca de 3,6 milhões de pessoas beneficiaram-se no Brasil da ação social do SESC [1]. Esse contingente é composto, em sua grande maioria, pelos trabalhadores do comércio de bens e serviços, seus familiares e dependentes. No entanto, o público atendido pelo SESC é muito maior, abrangendo também as populações da periferia de cidades de pequeno, médio e grande porte, que são assistidas pela entidade através de

parcerias com o poder público, empresas privadas, sindicatos e associações de moradores.

Vários dos serviços oferecidos são gratuitos. Para terem direito aos serviços, os associados devem apenas fazer a matrícula no SESC. No entanto, a maior parte dos associados, após a matrícula, nunca faz uso dos serviços. Tratar deste problema, conhecido como a não ativação do usuário, é o objetivo deste trabalho que propõe o desenvolvimento de um sistema de apoio à decisão para estimar, no ato da matrícula, o risco de não ativação do associado e identificar as possíveis causas desse risco em função do perfil do associado. Isso possibilitaria a tomada de medidas, por parte do SESC, no sentido de ofertar atrativos especiais para cativar tais associados. Para desenvolver o sistema de apoio à decisão foram utilizadas técnicas cientificamente avançadas e uma amostra de dados dos associados matriculados nos centros de atividades do SESC de Pernambuco até 2003.

O artigo está organizado em cinco seções. A Seção 2 apresenta uma breve fundamentação teórica das redes *Multilayer Perceptron* (MLP) e do algoritmo de treinamento *Backpropagation*. A Seção 3 aborda o indutor de regras utilizado neste trabalho. A Seção 4 descreve a base de dados e os experimentos, avaliando os resultados alcançados pela rede neural e pelas regras extraídas. A Seção 5 tece os comentários finais sobre o trabalho desenvolvido, apresentando as conclusões, limitações e melhorias futuras.

2. Redes neurais

Para fazer a avaliação do nível de ativação dos usuários, optamos pelo modelo de rede neural *Multilayer Perceptron* (MLP) treinada com o algoritmo *Backpropagation* [3, 4], pois esse modelo tem sido utilizado com sucesso na classificação de padrões. Dentre as características mais atrativas das redes MLP é possível destacar a excelente capacidade de generalização, a simplicidade de operação, ser um aproximador universal de funções [5] e a geração de uma resposta contínua que permite uma decisão baseada em limiar.

A rede MLP é formada por conjuntos de unidades de processamento (usualmente chamados de neurônios artificiais) agrupadas em camadas de entrada, intermediárias (ou escondidas) e de saída. Em uma rede MLP típica estas camadas são dispostas em seqüência e os neurônios de uma camada são completamente conectados com os neurônios da camada antecessora. As conexões entre os neurônios são associadas a um peso. Os valores destes pesos são aplicados aos padrões de entrada da rede que determinam a magnitude dos estímulos apresentados a cada neurônio. Estes estímulos são usados como argumentos para as funções de ativação

contidas nos neurônios artificiais. Os resultados das funções de ativação são passados adiante até que, na última camada, o resultado final da rede neural é produzido. A função de ativação mais comumente encontrada em redes MLP é a função sigmóide.

O treinamento de uma rede neural com o algoritmo *Backpropagation* é realizado em dois passos. No primeiro, um padrão de entrada é apresentado para a camada de entrada e as unidades de processamento computam, camada por camada, os resultados das funções de ativação até que um conjunto de estímulos é finalmente fornecido à camada de saída que o utiliza para determinar a saída da rede. O segundo passo começa com o cálculo do erro observado entre as saídas da rede e as saídas desejadas. Os erros são então propagados retroativamente a partir da camada de saída até a camada de entrada e os pesos das conexões entre as camadas são ajustados utilizando uma taxa de aprendizagem em função do gradiente da função objetivo. Este processo ocorre de tal forma que é minimizada a diferença entre as saídas calculadas pela rede e as saídas desejadas. Os dois passos descritos formam um ciclo de treinamento. Normalmente vários ciclos são necessários para que os pesos da rede modelem adequadamente os dados de treinamento.

3. Indução de regras

Diversas técnicas de indução de regras poderiam ter sido utilizadas neste trabalho. Uma variação do algoritmo *A Priori* [6] conjugada com a técnica de refinamento de Major e Mangano [7] foi escolhida para ser investigada. Neste caso, a indução das regras é realizada em 3 etapas.

Na primeira etapa foi aplicado o algoritmo *A Priori* que consiste em selecionar os subconjuntos mais freqüentes a partir dos subconjuntos de mais baixa ordem. Assim, são induzidas as regras que satisfazem a uma cobertura mínima e que têm uma complexidade limitada.

A segunda etapa consiste em, automaticamente, selecionar as regras que atendem a níveis de cobertura mais amplos que os induzidos originalmente e a níveis de confiança de interesse.

A terceira etapa consiste em selecionar, semi-automaticamente, as regras mais interessantes sob o ponto de vista do especialista e através da eliminação de redundâncias. Ao especialista, além da qualidade técnica, interessam aspectos de custo, velocidade e facilidade de obtenção das informações. A qualidade das regras pode ser avaliada de duas formas: uma medida em termos do *lift* da regra (aumento da confiança da regra em relação à confiança média da amostra analisada) e do aumento da cobertura de uma regra pela redução da sua complexidade funcional (e.g. passar de 3 variáveis para

2 variáveis) medida em termos de correlação, informação mútua ou outra métrica de similaridade entre grandezas não determinísticas.

4. Investigação experimental

4.1. Domínio do problema

Para cumprir com maior eficiência o seu papel social, o SESC visa a prestar serviços à maior quantidade possível de cidadãos elegíveis (empregados do comércio, empregados de alguns ramos de atividade de serviços, seus familiares e dependentes).

Para ter direito de usufruto dos serviços, o usuário elegível precisa apenas se matricular no SESC. O problema é que a maior parte dos associados, mesmo tendo direito de uso dos serviços oferecidos, muitos dos quais gratuitos, nunca faz uso dos mesmos. Esta não ativação do associado reduz o alcance social dos investimentos realizados pelo SESC.

O objetivo deste trabalho foi desenvolver um sistema de apoio à decisão para, no ato da matrícula, estimar o risco de não ativação do usuário e identificar as possíveis causas. Nesta abordagem, a não ativação foi caracterizada pela não utilização de qualquer dos serviços oferecidos durante o primeiro ano após a matrícula do associado.

A solução proposta fornece a resposta em tempo real sugerindo ações a serem aplicadas pelo SESC na sua política de relacionamento com os associados.

4.2. Base de dados

A base de dados utilizada neste trabalho contém cerca de 45.000 usuários registrados nos centros de atividades do SESC de Pernambuco até 2003. A base original possui 42 variáveis das quais 14 foram consideradas irrelevantes por uma das seguintes razões: identificador único, redundância, mau preenchimento, confidencialidade, mapeamento para outras variáveis e variável a posteriori. As variáveis usadas na solução estão especificadas na Tabela 1.

Tabela 1. Descrição das variáveis

Nome	Descrição	Tipo
Variáveis dos usuários		
NIVEL_ESCOLAR_USU	Nível de escolaridade	C
CENTRO_TIT_USU	Centro de atividade do titular	C
FLAG_TITULAR	Indica se o usuário é titular ou dependente	C
SEXO_USU	Sexo	C
RENDA_USU	Renda	N
ESTADO_CIVIL_USU	Estado civil	C
MUNICIPIO_USU	Município	C
BAIRRO_USU	Bairro	C
CEP1_USU	Dígito mais significativo do CEP	C
CEP2_USU	2 dígitos mais significativos do	C

	CEP	
CEP3_USU	3 dígitos mais significativos do CEP	C
CEP4_USU	4 dígitos mais significativos do CEP	C
CENTRO_USU	Centro de atividade	C
CATEGORIA1_USU	Subcategoria	C
CATEGORIA2_USU	Categoria	C
AREA_SER	Área de atividade do serviço	C
MODALIDADE_SER	Modalidade da atividade do serviço	C
Variáveis da empresa		
RZ_SOCIAL_EMP	Razão social	C
BAIRRO_EMP	Bairro	C
MUNICIPIO_EMP	Município	C
ESTADO_EMP	Estado	C
REGIAO_EMP	Região	C
CEP1_EMP	Dígito mais significativo do CEP	C
CEP2_EMP	2 dígitos mais significativos do CEP	C
CEP3_EMP	3 dígitos mais significativos do CEP	C
CEP4_EMP	4 dígitos mais significativos do CEP	C
QTD_END_EMP	Quantidade de endereços	N
Saída ou Alvo		
STATUS	Indica se houve ativação	C

* Tipos: categórica(C) e numérica(N)

Cada variável CEP foi transformada em 4 variáveis contendo grupos de bits mais significativos. A variável *STATUS*, que possui a informação a posteriori, foi categorizada em uma variável dicotômica indicando se houve ou não ativação do usuário segundo os critérios estabelecidos pelos especialistas do SESC. Essa categorização resultou no alvo da função-objetivo da aprendizagem supervisionada, tanto para a estimação do risco de não ativação quanto para a indução das regras de classificação.

Após a seleção e tratamento das variáveis, dois processos de natureza e objetivos diferentes foram realizados: a codificação binária dos campos categóricos para compatibilizar com as entradas da rede neural e a discretização das variáveis numéricas em faixas para compatibilizar com o indutor de regras. Para a rede neural, os dados numéricos foram normalizados de maneira que assumissem valores entre 0 e 1. Dessa etapa resultaram duas visões da base de dados: uma para aplicação na rede neural e outra para o algoritmo de indução de regras.

A base de dados da rede neural foi dividida em 3 conjuntos (treinamento, validação e teste) com 50%, 25% e 25% dos registros, respectivamente [8]. Os conjuntos de treinamento e validação foram replicados a fim de igualar a quantidade de usuários ativados e não ativados. O conjunto de teste foi mantido com suas características originais. Os dados utilizados na indução das regras foram resultantes da junção dos registros do conjunto de treinamento (antes da replicação) com os registros do conjunto de validação.

4.3. Rede neural

4.3.1. Treinamento. Redes MLP foram utilizadas nos experimentos deste trabalho. Os pesos da rede foram inicializados com valores aleatórios próximos a zero (na ordem de 10^{-4}) e a função de ativação sigmóide foi aplicada nos neurônios. As redes neurais foram treinadas com o algoritmo *Backpropagation* com taxa de aprendizagem igual a 0,001 e momentum 0,01. O critério adotado para parada do treinamento foi o monitoramento do poder de generalização da rede no conjunto de validação (0,5), utilizando o MSE (*Mean Square Error*) como métrica [8]. Diversas combinações de taxas de aprendizado, momentum, pesos iniciais e poder de generalização foram experimentados antes da adoção dos valores citados.

A codificação das variáveis de entrada em valores numéricos expandiu o vetor de entrada para 141 dimensões, resultando em uma rede com 141 neurônios de entrada e 2 de saída. Na camada intermediária, foram utilizados 2 neurônios, de modo a oferecer à rede uma funcionalidade que preservasse o número de parâmetros $[(2+141) \times 2 = 286]$ inferior ao recomendado pela prática usual [8] para a quantidade de dados disponíveis no conjunto de treinamento (cerca de 22.000).

Considerando o critério de parada, a aprendizagem convergiu em 192 épocas. Para o conjunto de validação, o MSE obtido foi de 0,07819 e o $KS2_{Máx}$ de 78,50 (Seção 4.3.4). A Figura 1 mostra a evolução do MSE dos conjuntos de treinamento e validação ao longo das iterações.

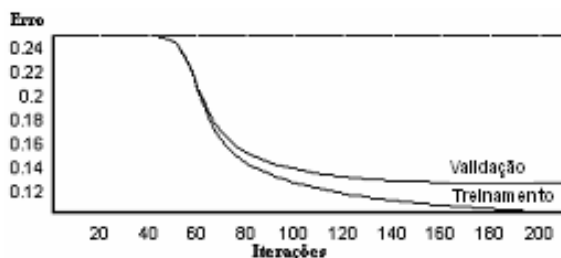


Figura 1. Evolução do MSE ao longo das iterações

4.3.2. Relevância das variáveis. A relevância das variáveis de entrada da rede foi analisada (Figura 2), o que possibilitou a observação de uma suavidade na curva. Portanto, uma solução de boa qualidade foi obtida uma vez que quase todas as variáveis contribuíram de significativamente para a tomada de decisão. As dez variáveis mais importantes são: *AREA_SERV*, *CATEGORIA2_USU*, *MODALIDADE_SERV*, *CATEGORIA1_USU*, *FLAG_TITULAR*, *MUNICIPIO_USU*, *CEP2_USU*, *SEXO_USU*, *CEP2_USU* e *CENTRO_USU*.

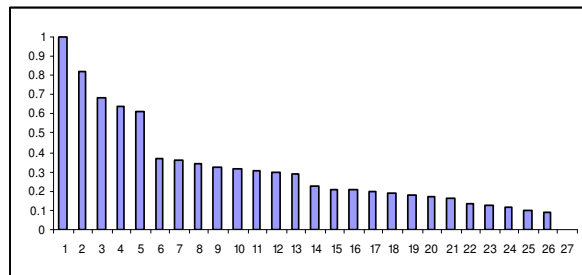


Figura 2. Relevância normalizada das entradas da rede neural.

4.3.3. Análise dos resultados. Considerando que o problema foi modelado como categórico (2 classes), a resposta de cada neurônio pode ser conjugada por meio de uma transformação linear em uma única grandeza escalar (score) definida no domínio contínuo de 0 a 100, onde 0 representa a não ativação e 100 significa a ativação do usuário. Essa representação da resposta por uma grandeza escalar possibilita um monitoramento muito mais refinado do desempenho do sistema de apoio à decisão.

A transformação da saída da rede em uma resposta contínua (score) demanda um limiar (ponto de corte) para definir a separação das classes. A escolha deste limiar depende de diversos fatores práticos tais como o custo relativo entre os dois tipos de erro (Tipo I e Tipo II) e o impacto da decisão. A métrica usada para medição do desempenho da rede foi o gráfico do teste estatístico de Kolmogorov-Smirnov (KS2) [9]. O teste KS2 é um teste não paramétrico geralmente utilizado para medir a aderência de dados a uma distribuição. Em sistemas decisórios, este teste serve para medir a separabilidade entre duas distribuições a partir da função de distribuição acumulada de cada uma. O KS2 é baseado na diferença das funções de distribuição acumulada, tendo como estatística do teste, o módulo da máxima diferença.

Os resultados apresentados nesta seção foram realizados sobre o conjunto de teste composto de 11.000 casos estatisticamente independentes dos dados de modelagem (treinamento e validação). Para medição dos resultados foram traçados os gráficos do KS2 e da distribuição das classes.

O gráfico da Figura 3 mostra que existe uma grande separação entre as distribuições dos escores dos usuários ativados e não ativados, o que indica que o sistema é capaz de prever a ativação com uma boa margem de acerto. O $KS2_{Máx}$ para o conjunto de teste foi de 70,89.

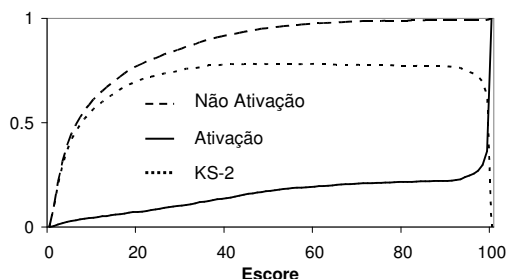


Figura 3. Funções da Distribuição Acumuladas (FDAs) e Curva de Kolmogorov-Smirnov (KS-2)

O gráfico da Figura 4 mostra que os usuários ativados estão concentrados, em sua maior parte, nos altos escores, ao passo que os usuários não ativados estão mais concentrados nos baixos escores.

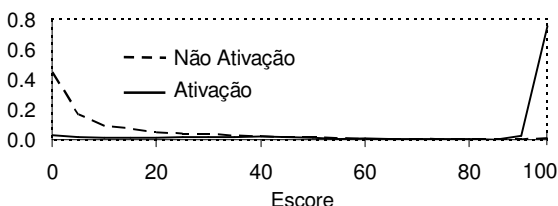


Figura 4. Distribuição de freqüência das classes em função do escore.

Neste sistema de apoio à decisão o ponto de corte (limiar de decisão) tem papel fundamental porque serve, por exemplo, para controlar o impacto das decisões em termos de carga de trabalho necessária, taxa de sucesso e retorno em benefício de cada ação tomada.

Tabela 2. Cenários de decisão

Limiar de decisão	% Usuários no Grupo-N	% Não ativados no Grupo-N	% Não ativados no Grupo-A
50	81,9%	96,3%	20,5%
60	83,9%	96,0%	12,8%
70	85,1%	95,7%	7,8%
80(*)	85,5%	95,6%	5,9%
90	85,7%	95,5%	5,3%

(*) Limiar de decisão sugerido

A Tabela 2 ilustra cenários de decisão com diversos valores do limiar de decisão e seus impactos nos serviços do SESC quanto à ativação dos usuários. A definição do ponto de operação do sistema cabe ao decisor humano, que se baseia nas políticas de assistência social em vigor no SESC para tomar esta decisão. Como neste estudo não dispomos de dados sobre custos e benefícios das decisões, os impactos ficarão restritos à concentração de usuários não ativados

no grupo selecionado para ações de ativação (Grupo-N) e no grupo não recomendado para ações de ativação (Grupo-A). Um aspecto importante para a definição do ponto de operação (limiar de decisão mais adequado à separação dos grupos) é a estabilização dos indicadores ao longo da variação do limiar (Tabela 2). Assim, o ponto de operação, definido em conjunto com os especialistas (em negrito), é o limiar mais à esquerda onde a concentração dos usuários fiéis atingiria máxima fidelidade; apenas 5,9% desse grupo, que representa 14,5% dos usuários, não atuariam.

4.4. Indução de regras

Após uma investigação experimental de diversas combinações de parâmetros do indutor de regras implementado, os seguintes parâmetros foram selecionados: suporte mínimo de 2% (é estatisticamente representativo e restringe o custo computacional a valores aceitáveis), dimensão máxima igual a 3 (limita a complexidade funcional e preserva um bom padrão de interpretabilidade das regras) e confiança (percentual de ativação) mínima de 92%. Dado que a confiança mínima é de 92% para os dois tipos de regras, o *lift* mínimo é igual a 1.11 para as regras de não ativação e 5.29 para as regras de ativação.

A Tabela 3, mostra 8 exemplos de regras que explicitam o conhecimento extraído sobre a ativação dos usuários. As quatro primeiras regras são referentes à não ativação dos usuários e as demais regras se referem à ativação.

Tabela 3. Exemplos de regras

Regra	Suporte	Confiança	Lift
Não ativação (Percentual de não ativação da massa inteira: 82.6%)			
CENTRO_TIT_USU = 193 E CENTRO_USU = Santa Rita	9.31%	97.82%	1.18
MUNICIPIO_USU = Olinda	8.15%	92.68%	1.12
CATEGORIA1_USU = Comerciário dependente	50.66%	93.86%	1.14
CENTRO_USU = Santa Rita E CATEGORIA2_USU = Com	8.31%	96.75%	1.17
Ativação (Percentual de ativação da massa inteira: 17.4%)			
AREA_SER = Educação Complementar	3.27%	100%	5.74
MODALIDADE_SER = Complementação Curricular	2.20%	100%	5.74
CATEGORIA2_USU = Usu	12.84%	92.85%	5.34
CATEGORIA1_USU = Usuário	12.58%	92.73%	5.33

Através de uma análise detalhada da base de dados e das regras extraídas, foi possível descobrir conhecimento interessante sobre o problema analisado. A Tabela 4 mostra alguns outros exemplos de nichos com maior e menor índice de ativação.

Tabela 4. Conhecimento descoberto.

Nichos de maior ativação
Centro de atividades de Petrolina
Cidades com menos opções de lazer
Centros de atividades que oferecem serviços diferenciados
Centros que oferecem cursos de complementação curricular
Categorias <i>usuário</i> e <i>usuário de 3ª. Idade</i>
Nichos de menor ativação
Usuários casados de Recife, Olinda e Jaboatão
Comerciários dependentes de Recife, Olinda e Jaboatão
Comerciários dependentes com 2º. grau completo
Comerciários dependentes solteiros
Centros de atividades de Santa Rita e Casa Amarela
Comerciário, comerciário dependente e comerciário aposentado
Associados de Gravata, Moreno, Belo Jardim, Floresta, Serra Talhada, São Lourenço da Mata, Igarassu, Camaragibe, Paulista, Cabo, Olinda, Recife e Abre e Lima
Centros de atividades distantes da residência ou do trabalho do associado

A partir do conhecimento descoberto é possível sugerir algumas medidas a serem tomadas pelo SESC para minimizar o problema de efetivação nos nichos de menor ativação. Alguns exemplos de sugestão são: oferecer serviços de boa aceitação por parte dos associados, como os de natureza de complementação curricular; oferecer serviços exclusivos; implementar um serviço de fidelização mais atrativo; questionar os associados, no momento da inscrição, sobre suas preferências em termos de serviços; considerar a questão geográfica (distância do trabalho ou residência para o centro de inscrição), a idade e o poder aquisitivo dos associados.

5. Considerações finais

Este trabalho apresentou uma aplicação de mineração de dados ao problema de ativação dos usuários dos serviços prestados pelo SESC. Os resultados alcançados foram estatisticamente relevantes com relação ao desempenho na estimação do nível de ativação dos usuários. Adicionalmente, as regras induzidas foram consideradas relevantes por um experiente especialista do SESC, pois permitiram um maior entendimento das principais causas que levam à não ativação dos usuários.

Na tabela 2 a linha em negrito ilustra o impacto para um ajuste específico do ponto de operação do sistema de apoio à decisão abaixo do qual todos os usuários devem ser estimulados a usar os serviços do SESC. Em vez dos 13,8% de ativação característicos da população cadastrada, nesse grupo que representa 14,5% da população, um total de 94,1% dos usuários farão uso dos serviços. Assim, estimular os demais usuários que representam 85,5% do total e têm um nível de ativação de apenas 4,4% é de importância fundamental para que o SESC aperfeiçoe o seu papel social.

A implantação do sistema de apoio à decisão possibilitaria análises bem mais aprimoradas das ações

sugeridas para cada perfil de usuário. Adicionalmente, o acompanhamento das decisões tomadas e dos seus impactos ao longo do tempo permitiria um refinamento do perfil dos usuários a partir de seu comportamento.

Apesar de este trabalho ter sido todo realizado com o acompanhamento de um especialista do SESC, algumas questões deixaram de ser tratadas e exploradas de forma mais detalhada em função da dificuldade de acesso ou inexistência de informação sistematizada.

Uma extensão natural deste trabalho seria a validação dos perfis de usuários de maior relevância social por meio de entrevista com representantes aleatoriamente amostrados nesses nichos.

O problema investigado neste trabalho é apenas um dos muitos de interesse para o setor público na prestação de serviços com fins sociais. Trabalhos desta natureza podem servir para empresas semelhantes ao SESC (SESI, SEST...) e empresas do Poder Público justificar para população uma grande parte das decisões tomadas e para prestar contas aos seus contribuintes de como os recursos arrecadados estão sendo utilizados.

6. Referências

- [1] Serviço Social do Comércio de Pernambuco — SESC-PE. www.sesc-pe.com.br.
- [2] Decreto-lei 9.853 de 13/09/1946.
- [3] Haykin, S., *Neural Networks: a comprehensive foundation*, Prentice Hall Press PTR, 2nd ed., 1998
- [4] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation", *MIT Press Computational Models of Cognition and Perception Series*, Vol. 1, 1986, pp. 318-362
- [5] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Network*, Vol. 2 (5), 1989, pp. 359-366
- [6] Witten, I.H., and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, California, USA, 2000
- [7] J. Major, and J. Mangano, "Selecting among Rules Induced from a Hurricane Database", In *Proc of 1993 KDD Workshop*, Washington, DC, July, 1993, pp. 28-47
- [8] L. Prechelt, "Proben1 – A Set of Neural Network Benchmark Problems and Benchmarking Rules", Technical Report 21/94, Fakultät für Informatik, Universität Karlsruhe, Germany, 1994
- [9] Conover, W.J., *Practical Nonparametric Statistics*, John Wiley & Sons, New York, 3rd ed., 1999, Chap. 6
- [10] Duda, R.O., P.E. Hart, and D.G. Stork, *Pattern classification*, Wiley-Interscience, 2nd ed., 2000