

Aplicação de Rede Bayesiana na identificação de fatores que causam a evasão no curso de Sistemas de Informação da UFRRJ.

Ramon Lameira¹, Robson Mariano da Silva¹

¹ Departamento de Matemática/Sistemas de Informação – Instituto de Ciências Exatas – Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Caixa Postal 15.064 – 91.501-970 – Seropédica, RJ – Brasil

ramon_lameira@hotmail.com, robsonms@ufrrj.br

Abstract. *We propose a computational model based on the use of Bayesian networks in order to identify factors leading to evasion in the undergraduate course in the Information Systems – UFRRJ. The proposed model was capable of identify the non-correlation of origin as school dropout factor. The sensitivity obtained in the forecast of evasion in the validation set was 93.3%. The results show promising regarding on the use of bayesian network, in identification of factors and prediction of student evasion.*

Resumo. *Propomos um modelo computacional baseado na utilização de redes bayesianas, a fim de identificar fatores que levam a evasão no curso de graduação em Sistemas de Informação da UFRRJ. O modelo proposto foi capaz de identificar a não correlação da origem escolar como fator de evasão. A sensibilidade obtida na predição de evasão no conjunto de validação foi de 93,3%. Os resultados obtidos mostram-se promissores quanto ao uso de rede bayesiana, na identificação de fatores e predição da evasão discente.*

1. Introdução

A evasão escolar é, certamente, um dos problemas que afligem as instituições de ensino em geral gerando prejuízos sociais, econômicos, e acadêmicos a todos os envolvidos no processo educacional. Segundo Martinho [2012], é um problema complexo advindo da superposição de fatores sociais, econômicos, culturais e acadêmicos, além de variáveis demográficas e atributos individuais que influenciam na decisão do estudante permanecer ou abandonar o curso.

Diante deste quadro, nos últimos anos, no Brasil, foram realizados monitoramentos, estudos empíricos, pesquisas científicas e levantamento estatístico sobre a evasão no ensino superior, com vistas à adoção de medidas visando mitigar os altos índices de evasão. Dentre as contribuições e ações resultantes desses processos podemos relacioná-las sob três paradigmas; 1- contribuições teóricas, que visam compreender e identificar os fatores que contribuem e influenciam no processo de evasão; 2 – ações específicas de cada instituição de ensino superior (IES), estruturada em sua realidade e especificidade; e, 3 – ações governamentais, como o Programa de Reestruturação e Expansão das universidades Federais (REUNI), [MEC/SESu, 2007], estabelecido, com as Instituições Federais de ensino Superior (IFEs).

Apesar do esforço dos órgãos governamentais, os estudos sobre evasão existentes no Brasil, ainda são incipientes e tímidos, se comparados com os países desenvolvidos onde os estudos e pesquisas sistematizados são inúmeros [Portela, 2013].

Neste contexto, é imprescindível realizar estudos que possibilite a identificação precoce de estudantes vulneráveis à evasão, de sorte a possibilitar a aplicação de ações proativas no sentido de evitar o abandono.

Nos últimos anos, diversas metodologias baseada em inteligência computacional, foram desenvolvidas visando avaliar a evasão escolar. Dentre essas metodologias, pode ser citado o trabalho de Mustafá [2012] que propõe a aplicação de árvore de regressão e classificação para identificar a evasão discente, tendo como base inicial os dados da inscrição do estudante em um curso presencial. Os resultados obtidos permitiram concluir que os dados utilizados conferem um baixo nível de precisão na identificação da evasão.

Martinho *et al.*, [2012] propuseram um sistema inteligente para a predição de grupos de risco de evasão discente nos cursos de tecnologia do Instituto Federal de Mato Grosso, utilizando rede neural ARTMAP-*Fuzzy*. Os resultados obtidos apresentaram índice de acerto do grupo evasivo de 97% e acerto global superior a 76%, valores estes que demonstram a confiabilidade, a acurácia e a eficiência do sistema.

O presente trabalho propõe um modelo computacional baseado em Redes Bayesianas para a identificação de fatores que causam a evasão discente no curso de graduação em Sistemas de Informação da UFRRJ.

Este artigo é estruturado da seguinte forma. Na Seção 1, a justificativa e o objetivo deste trabalho são descritos. A Seção 2 descreve os principais métodos e materiais utilizados ao longo do desenvolvimento deste trabalho. A Seção 3 mostra os resultados obtidos com a aplicação do modelo computacional em comparação com dados da literatura. A Seção 4, além de enfatizar as realizações que o trabalho conseguiu alcançar, também cita uma proposta de inserir novos parâmetros no modelo computacional.

2. Materiais e Métodos

O conjunto de dados constitui-se de 70% (n=105) dos alunos com matrícula no curso de Sistemas de Informação, obtidos junto a Pró-reitoria de Graduação da Universidade Federal Rural do Rio de Janeiro (UFRRJ). Sendo que desse universo $\approx 90\%$ (n=90) foram selecionados para compor o conjunto de treinamento do modelo e o conjunto de validação com $\approx 10\%$ (n=15) dos alunos com matrícula ativa ou não no referido curso escolhidos ao mero acaso. Com as seguintes variáveis: gênero (masculino ou feminino), origem escolar (pública ou privada), reprovação (sim ou não), distância (perto ou longe), para o parâmetro perto foi adotado uma distância inferior a 40 km da universidade e evasão (sim ou não).

O método proposto consiste na utilização de Redes Bayesianas, que é um modelo gráfico para relacionamentos probabilístico entre um conjunto de variáveis [Neapolitan, 2004]. Os sistemas estruturados em Rede Bayesiana são capazes de gerar automaticamente predições ou decisões mesmo na situação de inexistência de informações. Onde cada variável aleatória (VA) é representada por um nó da rede e cada nó de (VA) recebe conexões dos nós que têm influência direta sobre ele, além de possui uma tabela de probabilidade condicional, que quantifica a influência dos seus pais sobre ele. Cujas finalidade é a avaliar o impacto de cada variável na identificação de fatores que promovam a evasão e realizar a predição do grupo de estudantes, de

forma acurada e contínua, inferindo de maneira fidedigna. A estrutura da rede (Figura 1) foi estabelecida em função da disposição de dependências e independências entre as variáveis utilizadas. Para cada dependência entre os nós da rede, foi calculada a probabilidade *a priori* das variáveis utilizadas no modelo.

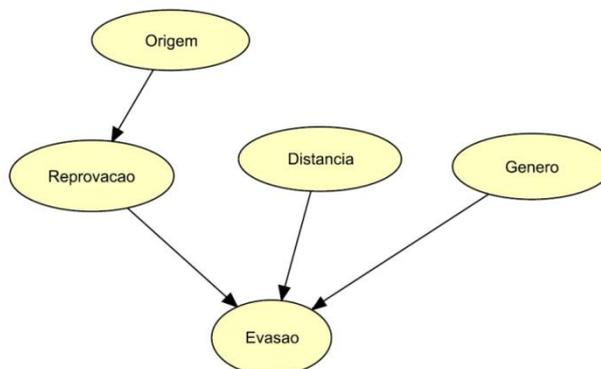


Figura 1. Fluxograma da Rede Bayesiana.

Para avaliar o desempenho do modelo neural bayesiano na predição de evasão no conjunto de teste, foram utilizados os índices de sensibilidade (verdadeiro positivo), especificidade (verdadeiro negativo). O modelo proposto foi implementado utilizando o pacote RHugin [Stefan e Achim, 2008] e o *software Hugin Expert* (www.hugin.com).

3. Resultados e Discussão

Neste trabalho foi desenvolvido uma rede bayesiana de sorte possibilitar a identificação de fatores que levam a evasão, bem como a sua predição. Os resultados obtidos pela inferência da rede implementada para o conjunto de dados em tela, em função das probabilidades propagadas, mostrou que não há evidência da contribuição da origem escolar na evasão dos discentes (Figura 2), fato este corroborado por [Nunes, 2013], que afirma que a origem não contribui para a evasão escolar. A taxa de evasão obtida foi de 72,3%, valor próximo ao descrito na literatura para a área de tecnologia da informação.

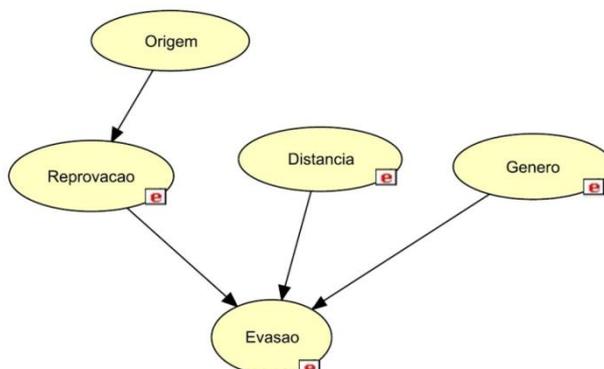


Figura 2. Dependências selecionadas pela rede bayesiana.

No que tange a variável reprovação o valor obtido pelo modelo foi de 89,1%, indicando uma forte correlação entre a reprovação e a evasão discente. Quanto à predição, o modelo computacional proposto obteve um erro de sensibilidade (verdadeiro positivo) para a evasão de 6,7%, e critério BIC -54,49% para o conjunto de validação (15 alunos).

4. Conclusão

A contribuição desse trabalho consistiu no desenvolvimento de um modelo computacional neural bayesiano na identificação e predição de evasão no curso de sistemas de Informação da UFRRJ.

Os resultados obtidos pelo modelo mostram-se promissores quanto à utilização da inferência bayesiana na identificação precoce dos fatores que levam a evasão no curso de Sistemas de Informação da UFRRJ. Além disso, o modelo foi capaz de identificar existência de correlação entre as variáveis e a evasão no referido curso.

Como proposta para trabalhos futuros pretende-se aplicar novas variáveis referentes a requisito didático/pedagógico, institucionais e sócio políticos/econômicos.

5. Referências

- Martinho, V.R.C.(2012) “Sistema Inteligente para Predição do Grupo de Risco de Evasão Discente”. Unesp, Faculdade de Engenharia. 58p.
- Martinho, V. R. C.; Nunes, C.; Minussi, C. R. (2012) “Um sistema inteligente usando redes neurais artmap-fuzzy para predição do grupo de risco de evasão discente em cursos superiores presenciais”. Simpósio Brasileiro de Automação Inteligente. Fortaleza, Brasil.
- MEC – Ministério da Educação (2007). Diretrizes Gerais do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais.
- Mustafa, M. N.; Chowdhury, L. and Kamal, S. (2012) “Students Dropout Prediction for Intelligent System from Tertiary Level in Developing Country”. International Conference on Informatics, Electronics & Vision. Dhaka, Bangladesh: IEEE
- Neapolitan, R.E. (2004)“Learning Bayesian Networks”. Upper Saddle River: Pearson.
- Nunes, R.C. (2013) “Panorama Geral da Evasão e Retenção no Ensino Superior no Brasil(IFES)”. XXVI Encontro Nacional de Pró-reitores de Graduação.
- Portela, S. (2013) “Evasão ou Retenção? Uma questão crucial à sustentabilidade das Instituições de Ensino Superior”, ABMESeduca.com, São Paulo. (acessado em 10.08.2014).
- Stefan, T., Achin, Z., (2008) “Collaborative software development usind R-Forge”. Report 81, Departament of Statistics and Mathematics, Wirtschaftsuniversitat Wien, Research Report Series.