

Rede Bayesiana para previsão de evasão escolar

Diogo Henrique Oliveira Souza diogo.souza@utp.edu.br

2 de dezembro de 2020

1 Introdução

Este artigo tem como propósito realizar a análise do artigo científico de inteligência artificial com ênfase na previsão de evasão escolar, produzido por Willian Silvano Maria, João Lucas Damiani e Max Roberto Pereira, onde apresenta a aplicação de Redes Bayesianas, com o intuito de prever os percentuais de chance de evasão dos alunos, com o objetivo de auxiliar os gestores educacionais na prevenção destes tipos de situações.

2 Fundamentação teórica

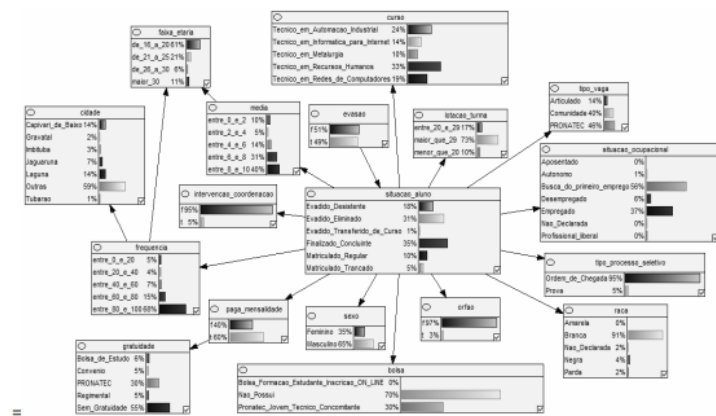
A predição foi realizada com base nas características dos alunos, coletadas do sistema utilizado pelo SENAI de Tubarão/SC, Para modelagem e testes da rede bayesiana foi utilizada a ferramenta Genie, uma ferramenta gratuita e livre de limitações. Foi utilizado, também no desenvolvimento, a API JSmile, responsável por realizar a inferência da rede modelada com o Genie que por padrão se utiliza do algoritmo de Lauritzen-Spiegelhalter, já a importação destes dados foi feita através do consumo de um webservice, que faz a leitura dos dados necessários no banco de dados do sistema SGN e os disponibilizam em formato de objetos JSON.

3 Modelagem da rede bayesiana

Para a construção do modelo da rede bayesiana proposta como solução, foi necessário efetuar um levantamento de quais seriam as principais informações que poderiam ser relacionadas aos alunos para posterior predição da evasão escolar. Inicialmente foi analisada a base de dados do Sistema de Gestão de Negócios (SGN), que é utilizado pela unidade do SENAI de Tubarão, a fim de encontrar informações importantes que pudessem ser adicionadas como nós

no modelo da rede bayesiana. Depois de uma análise dessas informações, foi elencado as possíveis variáveis candidatas para compor o modelo da rede bayesiana.

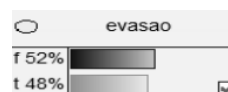
Figura 1: Modelo de rede bayesiana



(a) Fonte: www.finersistemas.com/atenaeditora/artigoPDF

As duas principais informações no modelo da rede são os nós "evasao" e "situacao-aluno", pois estas são as saídas da rede, ou seja, informações que o sistema disponibilizará, o primeiro nó caracteriza os percentuais de evasão dos alunos. Já o segundo nó caracteriza o percentual para cada situação em que o aluno poderá estar, A Figura 2 representa um exemplo do nó "evasao" na rede bayesiana. Podese observar que, de acordo com as informações dessa figura, o aluno teria 52% de probabilidade de não evadir e 48% de probabilidade de evadir do curso.

Figura 2: Nó que representa a evasão dos alunos.

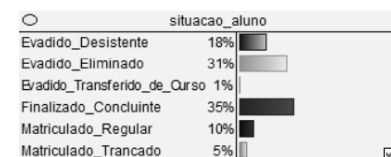


(a)

Fonte: www.finersistemas.com/atenaeditora/artigoPDF

A Figura 3 representa um exemplo do nó “situacao_aluno” na rede bayesiana. Pode-se observar que, de acordo com as informações dessa figura, o aluno teria 31% de probabilidade de evadir do curso com situação “Evadido_Eliminado”, 18% de evadir com situação “Evadido_Desistente”, 1% de evadir na situação “Evadido_Transferido_de_Curso”, 10% de se manter regular no curso, 5% de efetuar um trancamento e 35% de concluir o curso. As

Figura 3: Nó que representa as situações dos alunos.



(a)

Fonte: www.finersistemas.com/atenaeditora/artigoPDF

dependências condicionais entre os nós da rede podem ser verificadas na Figura 1, sendo estas representadas pelas setas que ligam os nós entre si. O nó de onde sai a seta é dependente do nó que recebe a seta. Desta forma pode-se verificar que o nó “evasao” é dependente do nó “situacao_aluno”, e que este, por sua vez, é dependente de vários outros nós, como por exemplo os nós: “sexo”, “media”, “frequencia”, dentre outros.

A Tabela a seguir representa a tabela de probabilidades do nó “situacao_aluno”, da rede bayesiana modelada, condicionada pelo nó “evasao”, desta forma pode-se dizer que o nó “situacao_aluno” é dependente do nó “evasao”, e a notação que representa essa dependência é $P(\text{situacao_aluno}|\text{evasao})$

Figura 4: Tabela de probabilidade condicional do nó “situacao_aluno”.

evasao	f	t
Evadido_Desistente	0	0.36625514
Evadido_Eliminado	0	0.63374486
Evadido_Transferido_de_Curso	0.01945525	0
Finalizado_Concluinte	0.6848249	0
Matriculado_Regular	0.19455253	0
Matriculado_Trancado	0.10116732	0

(a)

Fonte: www.finersistemas.com/atenaeditora/artigoPDF

3.1 Teorema de Bayes

O Teorema de Bayes é uma fórmula matemática usada para o cálculo da probabilidade

de um evento dado que outro evento já ocorreu, o que é chamado de probabilidade condicional, ele foi utilizado para efetuar o cálculo de probabilidades condicionadas onde existiam uma ou mais ligações entre os nós de evidências e hipótese. Como exemplo, no modelo de estudo, podemos citar os nós “frequencia” “media”, ambos dependentes do nó “faixa_etaria”. Nesse caso, para realizar os cálculos e preencher a tabela de probabilidades do nó “faixa_etaria”, foi aplicado o teorema de bayes, conforme representado na Equação 1. Na figura os “n” representam cada um dos estados que compõem os nós “faixa_etaria” (FE), “frequência” (FR) e “media” (MD).

Figura 5: Equação 1

$$P(FEn|FRn, MDn) = \frac{P(FEn).P(FRn|FEn).P(MDn|FEn)}{P(FRn, MDn)}$$

(a) Fonte: www.finersistemas.com/atenaeditora/artigoPDF

Após a conclusão da modelagem da rede bayesiana, tendo já definidos os nós e as dependências entre os mesmos, a etapa seguinte visa calcular e definir as tabelas das probabilidades condicionais e incondicionais da rede. É com base nessas tabelas que as inferências são realizadas sobre a rede, retornando os percentuais possíveis de evasão dos alunos nos cursos.

4 Análise do problema

O principal objetivo para os autores é de minimizar as chances de o aluno evadir, assim consequentemente auxiliando os gestores educacionais na prevenção destes tipos de situações, já que os números da evasão escolar no país têm sido um fator preocupante para as instituições de ensino, gestores e governantes, chegando a um percentual de 24,3% no ano de 2012 (Paim, 2013).

As principais dificuldades encontram-se na própria modularização do problema quando falamos em escalas maiores, já que instituições de ensino no geral, possuem números elevados de alunos, assim tornando a aplicação de redes bayesianas mais complexas, tendo a necessidade de inclusão de novos nós, estados e relacionar as dependências entre os nós da rede, além de ser necessário a frequente atualização dos dados de

coleta aos quais estão em constante mudança, visando alcançar a maior taxa de acerto possível.

5 Resultados experimentais

A aplicação da rede bayesiana foi implementada no sistema SPEED (Sistema Preditivo de Evasão Escolar Discente) que serve de auxílio aos gestores na exposição da situação potencial de evasão de alunos nos cursos aos quais estão vinculados e desta forma servindo de monitoramento para posteriores medidas pedagógicas a serem tomadas por estes gestores.

O sistema carrega todas as variáveis do aluno em questão que compõem o modelo da rede. Algumas das informações são fixadas, sem possibilidade de simulação, como por exemplo: o sexo, a idade, raça, etc. Outras informações ficam habilitadas para que o gestor possa efetuar simulações, como por exemplo, a média, a frequência, tipo de vaga, dentre outras informações. Ao selecionar as informações que irão compor o cenário de simulação, o sistema define como verdade cada uma das informações que o gestor selecionou e retorna os percentuais para os nós “evasao” e “situacao_aluno”. Desta forma o gestor poderá comparar as características que o aluno possui atualmente com as características que foram simuladas a fim de verificar se com a simulação houve melhora e assim agir nos pontos necessários.

Figura 6: Tela de simulação de evasão dos alunos no sistema SPEED.



(a) Fonte: www.finersistemas.com/atenaeditora/artigoPDF

Para validação da eficiência da rede bayesiana implementada, foram feitos alguns testes. Há dois tipos de cálculos realizados por uma rede bayesiana: a atualização de crenças e a revisão de crenças. A atualização é o cálculo de

probabilidades das variáveis aleatórias e a revisão refere-se à obtenção das probabilidades das hipóteses diagnósticas (nó de saída) e a identificação da hipótese diagnóstica com maior valor de probabilidade, já a revisão é justamente identificar o motivo do percentual de erro gerado pela rede bayesiana, nesta análise em específico foram verificados de forma minuciosa os “acompanhamentos pedagógicos” (registros acadêmicos feitos pela coordenação do curso) dos alunos selecionados nesta amostra aleatória, além de uma análise do desempenho até o momento da evasão (nos casos em que foram concretizadas). Essas informações foram analisadas para precisar o real motivo da evasão e avaliar se foi um erro crítico cometido pela rede bayesiana ou se foram fatos atípicos não previstos pela rede.

O resultado final do desempenho da rede bayesiana modelada ao final do experimento obteve uma taxa de 85,6% de acerto e 14,4% de erro. Para os casos em que a rede errou a inferência, em comparação com a real saída, foi realizada uma análise detalhada, averiguando os motivos para este comportamento na rede bayesiana modelada.

6 Considerações finais

A evasão é um problema mundial, mas ainda traz muita dor de cabeça para as instituições de ensino, tendo várias consequências tanto para o ambiente educacional quanto social. O sistema desenvolvido alcançou uma considerável taxa de acerto nas predições de evasão com base na rede bayesiana modelada, o que é ótimo para um primeiro protótipo, consequentemente com a evolução do software será possível obter resultados ainda mais satisfatórios e em proporções maiores, onde será possível auxiliar na resolução desse tipo de problema no âmbito educacional.

Referências bibliográficas

MARTINS, R.E. *Informática Aplicada à Educação 2* disponível em <https://www.finersistemas.com/atenaeditora/index.php/admin/api/artigoPDF/11419>. Acesso em: Novembro, 2020.