

Aplicação de Técnicas de Mineração de Dados para o Mapeamento do Conhecimento na Aprendizagem de Programação: Uma Estratégia Baseada na Taxonomia de Bloom

Rozelma Soares de França^{1,2}, Haroldo José Costa do Amaral¹

¹Universidade de Pernambuco, Campus Garanhuns (UPE)
CEP 55.294-902 – Garanhuns – PE – Brasil

²Centro de Informática – Universidade Federal de Pernambuco (UFPE)
CEP 50.740-560 – Recife – PE – Brasil

{rozelma.soares, haroldo.amaral}@gmail.com

Abstract. *The continuous evaluation process of learning can generate a large amount of data, where computational techniques can be used in the discovery of knowledge, assisting educators in their activities. This paper proposes a set of attributes, based on Bloom's Taxonomy, and the application of mining techniques on data collected from assessments to perform the mapping of knowledge of students. Experiments were carried out in a programming discipline demonstrate the feasibility of the proposal, enabling the identification of the assimilation level of content and allowing pedagogical strategies to be properly planned.*

Resumo. *O processo contínuo de avaliação da aprendizagem pode gerar uma grande massa de dados, onde técnicas computacionais podem ser utilizadas na descoberta de informações, auxiliando educadores em suas atividades. Este trabalho propõe a definição de um conjunto de atributos, pautado na Taxonomia de Bloom, e a aplicação de técnicas de mineração em dados coletados a partir de avaliações, para realizar o mapeamento do conhecimento de estudantes. Experimentos realizados em uma disciplina de programação demonstram a viabilidade da proposta, possibilitando a identificação do nível de assimilação de conteúdos e permitindo que estratégias pedagógicas sejam adequadamente planejadas.*

1. Introdução

A avaliação é essencial à educação e indissociável enquanto concebida como problematização, questionamento e reflexão sobre a ação (HOFFMANN, 2010). Nessa perspectiva, ela desempenha um papel fundamental em promover a aprendizagem, produzindo informações que poderão ajudar estudantes e professores. Assim, a avaliação deixa de ser apenas um instrumento de verificação da aprendizagem e passa a atuar diretamente no processo de ensino e aprendizagem, permeando-o e auxiliando-o não mais como uma atividade em momentos estanques e pontuais (CERNY, 2001).

Há uma busca por uma avaliação formativa que pode ser entendida como uma prática de avaliação contínua, cujo objetivo principal é melhorar a aprendizagem em curso, contribuindo para o acompanhamento e orientação dos estudantes durante todo seu processo de formação (PERRENOUD, 1999). Essa modalidade de avaliação, segundo Hadji (2001), possui características informativas e reguladoras, ou seja, fornece informações aos dois atores do processo de ensino e aprendizagem: i) professor, que será informado dos efeitos

reais de suas ações, podendo regular sua ação pedagógica e ii) aprendiz, que terá a oportunidade de tomar consciência de suas dificuldades e, possivelmente, reconhecer e corrigir seus erros.

O processo de avaliação contínua da aprendizagem pode produzir uma grande massa de dados, no seu decorrer, requerendo procedimentos automáticos ou semi-automáticos para o tratamento e análise desses dados. Nesse contexto, pesquisas vêm sendo realizadas, investigando possibilidades de exploração dessas informações, através da utilização de técnicas de mineração de dados (ROMERO & VENTURA, 2007). Espera-se, com isso, que as possíveis descobertas realizadas tenham o potencial de auxiliar na compreensão de como os estudantes aprendem e identificar os principais fatores que causam impactos no processo de aprendizagem (GOTTARDO et al., 2012). Na Mineração de Dados Educacionais – *Educational Data Mining* (EDM) –, autores citam algumas possibilidades de pesquisa na área, a exemplo de: i) modelos para identificar alunos com dificuldades de aprendizagem e ii) meios de melhorar a qualidade do material didático (BAKER et al., 2011).

No que tange à aplicação de técnicas de mineração em dados obtidos a partir de avaliações da aprendizagem, pode-se mencionar as pesquisas de Pimentel et al. (2006) e Santos et al. (2012). No entanto, ressalta-se que, nos estudos desses autores, não se observa, durante a seleção dos dados, a inclusão de atributos que façam referência aos objetivos pedagógicos planejados a serem alcançados pelos aprendizes.

Assim, este artigo, fruto de um Trabalho de Conclusão de Curso da Licenciatura em Computação, tem como objetivo apresentar um conjunto de atributos, elaborado a partir de uma taxonomia com fins educacionais, a Taxonomia de Bloom (BLOOM et al., 1956; ANDERSON & KRATHWOHL, 2001), para aplicação de técnicas de mineração de dados em avaliações contínuas da aprendizagem, que viabilize realizar o mapeamento do conhecimento dos estudantes. Com isso, torna-se possível identificar quais conteúdos e em que níveis eles foram assimilados pelos aprendizes, permitindo, assim, que estratégias pedagógicas sejam adequadamente planejadas para grupos de aprendizes que apresentem as dificuldades identificadas.

Para validar a proposta apresentada, foi realizado um estudo de caso com a disciplina Programação Orientada a Objetos, do curso de Licenciatura em Computação da Universidade de Pernambuco. Esta escolha deu-se pelo fato de que tal disciplina tem apresentado altas taxas de reprovação, alcançando em 2010 uma taxa superior a 75%. Tal realidade não é diferente em outros cursos de graduação, na área de Computação, sendo considerado um dos gargalos existentes, dificultando ou até mesmo impedindo a continuidade dos alunos em seus estudos (RAPKIEWICZ et al., 2006).

O restante do trabalho está organizado como segue: a Seção 2 apresenta uma taxonomia de objetivos educacionais, tida como referencial pedagógico para a elaboração do conjunto de atributos proposto. A Seção 3 apresenta conceitos sobre mineração de dados, algoritmos para extração de padrões e contextualiza o uso dessas técnicas no âmbito da educação. As Seções 4 e 5 descrevem, respectivamente, um estudo de caso com a aplicação de técnicas de mineração, em dados obtidos a partir de sessões de avaliação da aprendizagem, e os resultados gerados por tal tarefa. Por fim, na Seção 6, são feitas algumas considerações acerca deste trabalho e os aprofundamentos necessários.

2. A Taxonomia de Bloom

A clara e estruturada definição de objetivos educacionais pode direcionar o processo de ensino para a escolha adequada de estratégias, métodos, delimitação do conteúdo específico, instrumentos de avaliação e, conseqüentemente, para uma aprendizagem efetiva e duradoura

(FERRAZ et al., 2010). Nesse contexto, um dos instrumentos existentes que pode vir a facilitar esse processo é a taxonomia proposta por Bloom et al. (1956), conhecida como Taxonomia de Bloom, que tem como objetivo, ajudar no planejamento, organização e controle das metas de aprendizagem.

A Taxonomia de Bloom é composta por três domínios: Cognitivo, Afetivo, Psicomotor; entretanto, o domínio cognitivo é o mais conhecido e utilizado. No domínio cognitivo, os objetivos foram agrupados em seis diferentes categorias e são apresentados numa hierarquia de complexidade. As categorias desse domínio são: *Conhecimento*, *Compreensão*, *Aplicação*, *Análise*, *Síntese* e *Avaliação*. Com a incorporação de novos conceitos, recursos e teorias ao campo educacional, bem como os avanços psicopedagógicos e tecnológicos, e a publicação de diversas experiências de sucesso, no uso efetivo da taxonomia, observou-se a necessidade de reavaliação e releitura dos pressupostos teóricos que sustentam a pesquisa original apresentada em 1956 (BLOOM et al., 1956) e, assim, a Taxonomia de Bloom foi revisada e atualizada (ANDERSON & KRATHWOHL, 2001).

Na taxonomia revisada, a base original das categorias foi mantida; todavia, ao separar, conceitualmente, o conhecimento do processo cognitivo, ocorreram as seguintes mudanças (KRATHWOHL, 2002): i) os aspectos verbais utilizados na categoria Conhecimento foram mantidos, mas essa foi renomeada para *Lembrar*; ii) Compreensão foi renomeada para *Entender*; iii) Aplicação, Análise e Avaliação, foram alteradas para a forma verbal *Aplicar*, *Analisar* e *Avaliar*, respectivamente, por expressarem melhor a ação pretendida e serem condizentes com o que se espera de resultado a determinado estímulo de instrução; iv) Síntese trocou de lugar com Avaliação (*Avaliar*) e foi rebatizada para a forma verbal *Criar*; v) os nomes das subcategorias existentes foram alterados para verbos no gerúndio.

Apesar de manter o *design* hierárquico da original, a nova taxonomia é flexível, pois, como aponta Krathwohl (2002), considerou a possibilidade de interpolação das categorias do processo cognitivo, quando necessário, devido ao fato de que determinados conteúdos podem ser mais fáceis de serem assimilados a partir do estímulo pertencente a uma mais complexa. Por exemplo, pode ser mais fácil entender um assunto após aplicá-lo e só então ser capaz de explicá-lo.

3. Descoberta de Conhecimento em Bases de Dados

Com o avanço das tecnologias computacionais, que permitem o armazenamento e o processamento de um grande volume de dados, novas tecnologias têm sido desenvolvidas para auxiliar na extração de informações dessas bases de dados, através de técnicas, como a Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Database – KDD*) e a Mineração de Dados (*Data Mining – DM*).

Segundo Fayyad (1996), a KDD consiste num processo de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis. Por outro lado, o processo de DM restringe-se à aplicação de algoritmos para extração de padrões (FAYYAD et al., 2002).

Este trabalho debruça-se sobre a aplicação da KDD, na tentativa de descobrir padrões em dados de avaliações da aprendizagem, com foco na etapa de DM. Nesse sentido, o processo da KDD é classificado em três grandes fases, como ilustra a Subseção 3.1, tendo em vista as abordagens descritas em Rezende et al. (2003) e Pimentel et al. (2006): Preparação, Extração de Padrões e Pós-Processamento.

3.1. Fases do KDD

Na fase de **Preparação**, ocorre a identificação e o entendimento do problema, considerando aspectos como objetivos e as fontes de dados das quais se pretende extrair o conhecimento. O passo seguinte consiste na seleção dos dados, a partir das fontes, de acordo com os objetivos do processo, e no tratamento dos dados para poderem ser submetidos aos métodos e ferramentas, na fase de extração de padrões. Na fase de **Extração de Padrões**, escolhe-se inicialmente um ou mais algoritmos para a extração do conhecimento, podendo ser executada várias vezes para o ajuste de parâmetros, para obter resultados adequados aos objetivos estabelecidos. Na fase de **Pós-Processamento**, o conhecimento extraído é avaliado quanto à sua qualidade e/ou utilidade para que possa ser utilizado a fim de apoiar algum processo de tomada de decisão, seja por um especialista humano ou por um sistema especialista.

3.2. Tarefas e Técnicas de Mineração de Dados

No processo da KDD, a mineração de dados é a etapa central que executa a descoberta de conhecimento propriamente dita, pois são os seus algoritmos que produzem, de maneira semi-automática, o conhecimento a partir de dados existentes. Segundo Pimentel et al. (2006), as tarefas de mineração de dados são as classes de problemas determinadas de acordo com o tipo de conhecimento a ser minerado e, também, objetivos almejavéis para a solução. A escolha da técnica de mineração e do algoritmo a ser utilizado, contudo, depende da tarefa que se quer executar.

Nas Subseções 3.2.1 e 3.2.2, são descritas as tarefas e técnicas de mineração de dados que foram utilizadas no estudo de caso, descrito nas Seções 4 e 5, deste artigo. Para este trabalho, considera-se que o uso dessas tarefas é complementar e não competitivo. Ainda, recomenda-se o uso de Agrupamento, tarefa não descrita neste artigo por limitações de espaço.

3.2.1 Classificação

Conforme Han & Kamber (2001), a classificação de dados consiste no processo de predizer a classe de um objeto segundo certos atributos de predição. Uma das técnicas utilizadas é a *Árvore de Decisão (Decision Tree)* que compreende em escolher uma variável que se quer avaliar e o algoritmo procura as características mais fortemente relacionadas a ela, montando uma árvore de ramificações na forma SE-ENTÃO. O *J4.8* é um dos algoritmos que pode implementar essa técnica (WITTEN et al., 2011), sendo uma implementação do algoritmo *C.45*, proposto por Quinlan (1993), para geração *top-down* de árvores de decisão. Segundo, Witten et al. (2011), o *J4.8* utiliza uma técnica de busca gulosa, determinando a cada passo qual o atributo, dentre os disponíveis, é mais preditivo, e dividindo um nodo da árvore com base neste atributo.

3.2.2 Associação

Segundo Carvalho (2000), uma regra de associação é definida como: se X então Y ou $X \rightarrow Y$, onde X e Y são conjuntos de itens no formato atributo-valor e $X \cap Y = \emptyset$. Diz-se que X é o antecedente da regra e Y é o conseqüente da regra. A meta é descobrir se X implica Y . A técnica de regra de associação implementa esse tipo de tarefa através de algoritmos como o *Apriori*. De acordo com Witten & Frank (2005), esse algoritmo, na ferramenta WEKA, possui as seguintes características: o valor de suporte mínimo começa com 100% e vai reduzindo 5% até que 10 regras sejam geradas com confiança mínima de 90% ou até que o valor de suporte atinja a um limite inferior a 10%. Esses valores padrões podem ser alterados.

3.3. Mineração de Dados Educacionais

Técnicas de mineração de dados podem ser aplicadas a uma variedade de contextos de tomada de decisão como *marketing*, finanças, manufatura, saúde. Analogamente, é possível minerar dados de alunos para verificar a relação entre uma abordagem pedagógica e o seu aprendizado e, através dessa informação, o professor poder compreender se sua abordagem está realmente ajudando o estudante e desenvolver métodos de ensino mais eficazes (BAKER et al., 2011).

Em sistemas educacionais, segundo Romero & Ventura (2007), a descoberta do conhecimento pode ser orientada a diferentes atores de acordo com os interesses particulares de cada um: i) *Orientada para estudantes*: técnicas de mineração de dados são utilizadas para recomendar atividades, recursos e tarefas que favoreçam e melhorem a aprendizagem do aluno. Isso pode ser feito por meio de sugestões de *links*, baseadas nas experiências de aprendizagem de outros estudantes, por exemplo; ii) *Orientada para educadores*: objetiva fornecer *feedback* aos instrutores, avaliar a estrutura do conteúdo do curso e a sua eficácia no processo de aprendizagem. Também é foco de interesse classificar grupos de estudantes baseado em suas necessidades, descobrir padrões de aprendizagem regulares e irregulares, descobrir os erros frequentes, descobrir atividades que são mais eficazes, organizar o conteúdo de forma mais eficiente para o progresso do aluno; iii) *Orientada para gestores educacionais*: o objetivo é ter parâmetros sobre como melhorar a eficiência do sistema e adaptá-lo ao comportamento de seus usuários, ter medidas sobre como melhor organizar os recursos institucionais (humanos e materiais), melhorar a oferta de programas educacionais, entre outros.

4. Estudo de Caso

Visando à obtenção de um nível mais profundo de compreensão do conhecimento dos estudantes e suas dificuldades de aprendizagem, a fim de obter subsídios para superar essa realidade, um estudo de caso foi realizado, na disciplina Programação Orientada a Objetos, ministrada no primeiro semestre de 2010, onde havia 33 estudantes matriculados.

A abordagem adotada na disciplina foi além da avaliação convencional, baseada no rendimento alcançado pelo aluno em poucas provas. Adotaram-se, assim, avaliações formativas, ocorridas ao longo do processo de aprendizagem. Dentro desse escopo, foram considerados os dados obtidos a partir de 11 sessões de avaliação, a saber: 5 listas de exercícios, 4 miniprovas e 2 projetos. As atividades ocorreram durante todo o semestre e contemplaram diferentes conteúdos. Cada sessão era composta por “n” enunciados de problemas, totalizando 890 exercícios a serem avaliados pelo professor. Todas essas atividades antecederam às duas avaliações somativas, ocorridas ao fim de cada bimestre, e a análise e interpretação desses dados podem ser exploradas a fim de fornecer ao docente um melhor discernimento sobre o conhecimento e dificuldades dos estudantes na aprendizagem. As avaliações foram formuladas de modo a obter uma medida cognitiva, que deveria retratar o real desempenho do estudante na resolução de cada problema e seria resultado do processo de correção do professor.

Após a análise dos enunciados dos problemas, foi possível classificá-los de acordo com os níveis da Taxonomia de Bloom. Pelo fato de as questões não terem sido formuladas, tendo como referência uma taxonomia de objetivos educacionais, uma limitação deste trabalho está na classificação, já que o professor-autor dessas tarefas poderia ter um objetivo diferente do apresentado nos enunciados e ter equivocadamente utilizado um verbete de um

nível cognitivo que não desejava avaliar naquele momento nos estudantes, levando, assim, à classificação de uma questão em um nível superior ou inferior ao desejado pelo docente.

Ao final de cada bimestre, após a realização de avaliações formativas, pelos 33 estudantes matriculados na disciplina, foram realizadas avaliações somativas sendo que os valores possíveis estavam na faixa entre 0 e 10. Para efeitos de análise, os estudantes foram agrupados em dois grupos distintos: os que alcançaram uma média igual ou superior a 7, média requerida para aprovação, foram rotulados como aprovados, e os demais, que obtiveram uma média inferior a 7 ou um número de faltas superior a 25%, como reprovados.

A análise das médias das avaliações somativas revelou uma taxa de sucesso de apenas 24,24% entre os estudantes da disciplina. Tal taxa de sucesso demonstra claramente a dificuldade encontrada por esses estudantes. A partir dessa realidade, passou-se a tentar identificar se, dentre os dados disponíveis com o histórico das avaliações formativas, haveria evidências que pudessem identificar os fatores que dificultaram a aprendizagem na disciplina.

Para aplicação das técnicas de mineração, após o conhecimento do domínio, partiu-se para a etapa de preparação dos dados. Esta envolveu a seleção, o pré-processamento e limpeza e a transformação dos dados de maneira mais apropriada para a mineração. O Quadro 1 apresenta o conjunto de atributos proposto para este estudo de caso, junto com as suas descrições, tipos e domínio de valores que eles podem assumir.

Quadro 1. Atributos propostos para o mapeamento do conhecimento do estudante

Atributo	Descrição	Tipo de Dado	Domínio
IdAluno	Código que identifica o aluno participante da sessão de avaliação	Numeric	[1, n]
IdSessao	Código que identifica o número da sessão de avaliação e a sequência temporal	Numeric	[1, n]
IdItem	Código que identifica uma questão numa avaliação formativa	Nominal	[Q ₀₁ , Q _n]
Conteudo	Conteúdo tratado no item avaliado	Nominal	[C ₀₁ , C _n]
NivelCognitivo	Nível cognitivo do item avaliado de acordo com a Taxonomia de Bloom	Nominal	LEM – Lembrar ENT – Entender APL – Aplicar ANA – Analisar AVA – Avaliar CRI – Criar
ConceitoItem	Grau de acerto do aprendiz no item avaliado	Nominal	A – 8 ≥ nota ≤ 10 B – 6 ≥ nota < 8 C – 4 ≥ nota < 6 D – 2 ≥ nota < 4 E – 0 ≥ nota < 2 ? – Ausência de informação
ConceitoAvaliacao	Nível de desempenho do aprendiz numa determinada avaliação formativa	Nominal	A – 8 ≥ nota ≤ 10 B – 6 ≥ nota < 8 C – 4 ≥ nota < 6 D – 2 ≥ nota < 4 E – 0 ≥ nota < 2 ? – Ausência de informação
RotuloSomativa	Rótulo da avaliação somativa	Nominal	APR – Aprovado REP – Reprovado

A partir do conjunto de atributos, apresentado no Quadro 1, algumas questões podem ser formuladas na tentativa de descobrir conhecimento através dos algoritmos de reconhecimento de padrões. Eis alguns exemplos:

- Existe relação entre os níveis cognitivos da Taxonomia de Bloom e o desempenho do aprendiz nas avaliações?

- Existe relação entre o desempenho do aprendiz e a sequência cognitiva com que um conteúdo é trabalhado, como, por exemplo: *Os estudantes apresentam melhor desempenho, em determinado conteúdo, quando o mesmo é trabalhado primeiramente no nível Aplicar e posteriormente no nível Entender?*
- Existe relação entre o desempenho do aprendiz nas avaliações formativas e o rendimento na avaliação somativa?

5. Extração de Padrões e Análise de Resultados

Para aplicar as técnicas de mineração de dados, utilizou-se neste estudo de caso a ferramenta WEKA, desenvolvida pela Universidade de Waikato da Nova Zelândia (WITTEN & FRANK, 2005). A seguir, são descritas as duas tarefas de mineração de dados utilizadas neste trabalho: classificação e associação. A escolha dos algoritmos que implementam as técnicas utilizadas, nas tarefas mencionadas, baseia-se nos casos de sucesso apresentados por trabalhos correlatos, destacados na Seção 1, que demonstraram que os algoritmos *J4.8* e *Apriori* obtiveram um bom resultado.

5.1. Tarefa de Classificação

Para esta tarefa, foi necessário escolher um atributo alvo e os atributos de predição. Assim, escolheu-se como atributo alvo o conceito obtido pelo aprendiz na avaliação, ou seja, o nível de desempenho do estudante numa determinada avaliação formativa: A, B, C, D, E ou ?. A questão por trás desta escolha foi: *O que determina o desempenho do aprendiz numa avaliação formativa?*

O algoritmo de árvore de decisão escolhido foi o *J4.8*. Na Figura 1, é possível visualizar parte da árvore de decisão gerada com base nos 890 registros transformados a partir dos dados originais.

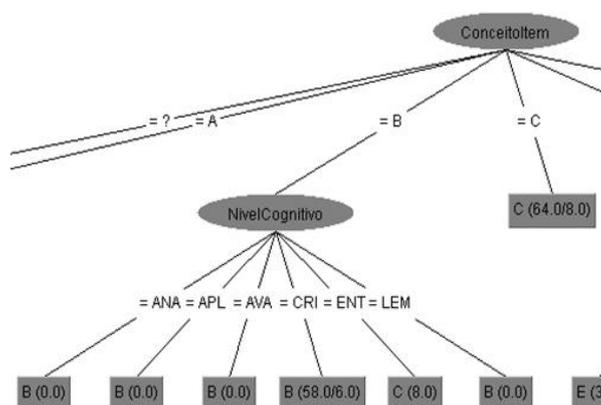


Figura 1. Parte da árvore de decisão gerada com o atributo alvo *ConceitoAvaliacao*

Analisando os padrões encontrados, podem-se realizar as seguintes interpretações:

- Percebe-se que se o conceito obtido no item for igual a “B” (nota variando entre 6,0 e 8,0), então o conceito final na avaliação formativa dependerá do nível cognitivo em que o item foi classificado:
 - Em 58 casos, exceto em 6, obtém-se conceito “B” na avaliação formativa se o item foi classificado no nível Criar;
 - Em 8 casos, obtém-se conceito “C” na avaliação formativa se o item foi classificado no nível Entender.
- Em 64 ocorrências, exceto em 8, quando o estudante obtém o conceito “C” (nota entre 4,0 e 6,0) no item, ele tem desempenho final “C” na avaliação formativa.

Pode-se concluir que, pela árvore de decisão, o nível cognitivo em que o item é classificado interfere no desempenho final do aprendiz na avaliação formativa.

Após a descrição da árvore de decisão, outro atributo foi selecionado para análise. Dessa vez, escolheu-se como atributo alvo o rótulo na avaliação somativa (APR - Aprovado ou REP - Reprovado) e procurou-se responder a seguinte questão: *Existe relação entre o desempenho do aprendiz nas avaliações formativas e o rendimento na avaliação somativa?*. A Figura 2 apresenta a nova árvore de decisão gerada.

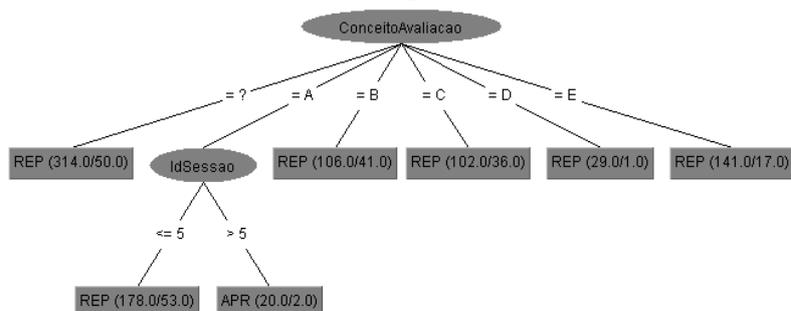


Figura 2. Árvore de decisão gerada com o atributo alvo *RotuloSomativa*

Analisando-se os padrões encontrados pelo algoritmo *J4.8*, representados na árvore de decisão da Figura 2, pode-se realizar as seguintes leituras:

- Percebe-se que, quando o estudante obtém o conceito “A” na avaliação formativa, o rótulo da avaliação somativa depende do ID da avaliação formativa, sendo:
 - Rótulo APR (aprovado) se o ID da sessão de avaliação for superior a 5;
 - Rótulo REP(reprovado) se o ID da sessão de avaliação for igual ou inferior a 5.
- Nota-se também que, se o aprendiz obtém conceito “?”, “B”, “C”, “D” ou “E”, na avaliação formativa, o rótulo que ele recebe na avaliação somativa é REP.

Pode-se concluir que, com base nesses resultados, o desempenho pleno nas avaliações formativas 6, 7, 8, 9, 10 e 11¹ é determinante para a aprovação do estudante. Outros atributos alvos podem ser selecionados para fins de análises comparativas.

5.2. Tarefa de Associação

A Figura 3 apresenta as 9 regras geradas pelo algoritmo de regras de associação *Apriori*.

1. NivelCognitivo=CRI ConceitoAvaliacao=? 210 ==> ConceitoItem=? 210 <conf:(1)>
2. NivelCognitivo=CRI ConceitoAvaliacao=? RotuloSomativa=REP 168 ==> ConceitoItem=? 168 <conf:(1)>
3. Conteudo=Estruturas de Repeticao 132 ==> NivelCognitivo=CRI 132 <conf:(1)>
4. Conteudo=Estruturas de Repeticao RotuloSomativa=REP 100 ==> NivelCognitivo=CRI 100 <conf:(1)>
5. ConceitoAvaliacao=? 314 ==> ConceitoItem=? 313 <conf:(1)>
6. ConceitoAvaliacao=? RotuloSomativa=REP 264 ==> ConceitoItem=? 263 <conf:(1)>
7. ConceitoAvaliacao=A RotuloSomativa=REP 127 ==> ConceitoItem=A 119 <conf:(0.94)>
8. Conteudo=Arrays ConceitoItem=? 107 ==> NivelCognitivo=CRI 99 <conf:(0.93)>
9. ConceitoAvaliacao=A 198 ==> ConceitoItem=A 180 <conf:(0.91)>

Figura 3. Regras de associação geradas pelo algoritmo *Apriori*

Analisando-se os padrões, as seguintes interpretações, dentre outras, podem ser feitas:

¹ Na disciplina Programação Orientada a Objetos ofertada no curso de Licenciatura em Computação da Universidade de Pernambuco em 2010.1, 11 avaliações formativas foram realizadas. Tais avaliações foram classificadas da seguinte forma: Lista de Exercícios (1, 3, 5, 8 e 10); Miniprovas (2, 4, 6 e 9); Projeto (7 e 11).

- Na regra 3, com confiança de 100%, tem-se 132 casos que, quando o conteúdo é Estruturas de Repetição, o nível cognitivo explorado é o Criar;
- Na regra 4, com 100% de confiança, percebe-se que, quando o conteúdo é Estruturas de Repetição e o aprendiz obtém rótulo Reprovado na avaliação somativa, a abordagem de ensino fez uso do nível Criar;
- Na regra 8, com confiança de 93%, tem-se que, quando o conteúdo é *Arrays* e não houve fornecimento de informação, pelo aprendiz, na resolução de um item que trate sobre esse conceito, o nível cognitivo empregado no questionamento foi o Criar.

6. Considerações Finais

O processo contínuo de avaliação da aprendizagem pode gerar uma grande massa de dados que necessita ser tratada e interpretada de modo a fornecer informações precisas acerca do conhecimento dos aprendizes, onde técnicas de mineração de dados podem ser utilizadas para tal fim, como demonstrado neste trabalho. Com a incorporação de referencial pedagógico, junto às técnicas computacionais, aspectos específicos do processo de ensino e aprendizagem podem ser descobertos para tomada de decisões.

Neste trabalho, buscou-se aporte teórico em Benjamin Bloom que, juntamente com outros pesquisadores, propuseram a Taxonomia de Bloom, uma taxonomia de objetivos educacionais que pode auxiliar os educadores em tarefas como planejamento e avaliação da aprendizagem. Os resultados obtidos demonstraram as potencialidades de técnicas de mineração de dados aliadas a objetivos educacionais. Em uma turma de Programação Orientada a Objetos, foi possível diagnosticar que existe relação entre os níveis cognitivos da Taxonomia de Bloom e o desempenho do aprendiz, como também entre o desempenho do aprendiz nas avaliações formativas e o rendimento nas somativas.

Através da análise de padrões de regras de associação, constatou-se que um dos gargalos na aprendizagem está no processo de criação (nível *Criar* da Taxonomia de Bloom), a exemplo de estruturas de repetição e *arrays*. Este fator atrelado a outros, identificados em atividades de monitorias, como: i) a incapacidade do estudante em estabelecer uma nova rotina de estudos que lhe permita fazer uso de uma melhor gerência de tempo e/ou ii) a desmotivação do aprendiz em aprender os novos conceitos, muitas vezes tidos como de difícil compreensão, tem influenciado no aumento dos altos índices de reprovação nessa disciplina. Na abordagem associativa, o uso do algoritmo *Apriori* gerou regras que parecem “óbvias”. Assim, sugere-se que outros algoritmos associativos sejam testados.

Além das tarefas descritas, outras podem ser utilizadas de modo complementar, a exemplo de Agrupamento. Através desta, torna-se possível agrupar estudantes pelas suas dificuldades de aprendizagem, possibilitando que estratégias pedagógicas sejam formuladas, a exemplo de atividades de monitoria dirigidas a grupos de aprendizes com tendência a não alcançar o desempenho requerido. Como trabalhos futuros, pretende-se aprofundar os testes e análise dos padrões encontrados, bem como replicar o método utilizado em outras turmas e avaliar a intervenção dos resultados. Somado a isso, planeja-se incorporar ao conjunto de atributos proposto outros dados que carreguem informações a despeito das estratégias utilizadas pelos estudantes na resolução das atividades propostas.

Referências

Anderson, L. W.; Krathwohl, D. R. (Ed.) (2001). “A taxonomy for learning, teaching, and assessing: A revision of Bloom’s taxonomy of educational objectives”. New York: Addison Wesley Longman.

- Baker, R.; Isotani, S.; Carvalho, A. (2011). "Mineração de Dados Educacionais: Oportunidades para o Brasil". *Revista Brasileira de Informática na Educação*, v. 19, n. 02, p. 3-13.
- Bloom, B.S. (Ed.); Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R.. (1956) "Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain". New York: David McKay.
- Carvalho, J. V. (2000). "Reconhecimento de Caracteres Manuscritos Utilizando Regras de Associação". Dissertação de Mestrado. UFCG/COPIN.
- Cerny, R. Z. (2001). "Uma reflexão sobre a avaliação formativa na educação a distância". UFSC.
- Fayyad, U.; Grinstein, G. G.; Wierse, A. (2002). "Information Visualization". In: *Data Mining And Knowledge Discovery*. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA.
- Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. (1996). "From Data Mining to Knowledge Discovery". American Association for Artificial Intelligence.
- Ferraz, A. P. do C. M.; Belhot, R. V.. (2010). "Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais". *Gest. Prod.*, São Carlos, v. 17, n. 2, p. 421-431.
- Gottardo, E.; Kaestner, C.; Noronha, R. V. (2012). "Previsão de Desempenho de Estudantes em Cursos EAD Utilizando Mineração de Dados: uma Estratégia Baseada em Séries Temporais". In: *XXIII Simpósio Brasileiro de Informática na Educação, 2012, Rio de Janeiro. Anais do XXIII SBIE*.
- Hadji, C. (2001). "A avaliação desmistificada". Porto Alegre: ArtMed.
- Han, J.; Kamber, M. (2001). "Data Mining: Concepts and Techniques". San Francisco: Morgan Kaufmann Publishers.
- Hoffmann, J. M. L. (2010). "Avaliação: mito e desafio – uma perspectiva construtivista". Porto Alegre: Mediação.
- Krathwohl, D. R. (2002). "A revision of Bloom's taxonomy: an overview. *Theory in Practice*", v. 41, n. 4, p. 212-218.
- Perrenoud, P. (1999). "Avaliação: da excelência à regulação das aprendizagens entre duas lógicas". Porto Alegre: Artes Médicas.
- Pimentel, E. P.; Omar, N. (2006). "Descobrendo Conhecimentos em Dados de Avaliação da Aprendizagem com Técnicas de Mineração de Dados". In: *XII Workshop de Informática na Escola, Campo Grande, MS. Anais do XXVI CSBC*.
- Quinlan, R. (1993). "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, San Mateo, CA.
- Rapkiewicz, C. E; Falkembach, G; Seixas, L.; Rosa, N. S.; Cunha, V. V. Da; Klemann, M. (2006). "Estratégias pedagógicas no ensino de algoritmos e programação associadas ao uso de jogos educacionais". *Revista Novas Tecnologias na Educação*, v. 4, nº 2, Dezembro.
- Rezende, S. O.; Pugliesi, J. B.; Melanda, E. A.; Paula, M. F. de. (2003). "Mineração de dados". In: *Rezende, S. O. Sistemas Inteligentes: Fundamentos e Aplicações*. Barueri, SP: Manole.
- Romero, C; Ventura, S.. "Educational Data Mining: A Survey from 1995 to 2005" (2007). *Expert Systems with Applications* 33, p. 125-146.
- Santos, H. L. dos; Camargo, F. N. P.; Camargo, S. da S. (2012). "Minerando Dados de Ambientes Virtuais de Aprendizagem para Predição de Desempenho de Estudantes". In: *Seventh Latin American Conference on Learning Objects and Technologies, Guayaquil. Proceedings of the 7th LACLO*.
- Witten, I. H.; Frank, E.; Hall, M. A. (2011). "Data Mining: Practical Machine Learning Tools and Techniques", 3rd edition, Morgan Kaufmann, Burlington, MA.
- Witten, I. H.; Frank, E. (2005). "Data Mining: Practical Machine Learning Tools and Techniques". 2nd edition, Morgan Kaufmann Publishers, San Francisco, CA.