Mineração de regras de associação para seleção de oferta de cursos de especialização

Marcos André S. Kutova **Clodoveu A. Davis Jr.**

Instituto de Informática – Pontifícia Universidade Católica de Minas Gerais
Av: Dom José Gaspar, 500 – 30535-901 – Belo Horizonte – MG – Brasil
{kutova,clodoveu}@pucminas.br

Abstract. *This paper presents association rule mining on data from a poll on the interest of undergraduate students on a range of graduate courses offered by a university. The goal was to uncover unexpected relationships among courses, to support the decision making process as to the potential competition among them. Generation of association rules allowed the authors to point out situations in which two or more courses share roughly the same group of interested students, and may therefore divide their potential public, possibly making each other economically unfeasible. Conclusions include recommendations on strategies for the seasonal offering of graduate courses.*

1. Introdução

Regras de associação identificam, com determinado nível de confiança, as relações entre dois ou mais itens de uma base de dados (Agrawal et al., 1993). Em geral, essas regras sugerem que a presença de um ou mais itens em um conjunto geralmente é acompanhada da presença de outro. Essas regras são interessantes porque normalmente estão escondidas em grandes bases de dados e, sem o auxílio de algum processamento computacional, pode ser difícil encontrá-las (Carvalho, 2005).

A facilidade com que as regras de associação podem ser construídas, com o apoio de algoritmos adequados, propicia a sua aplicação em outros contextos. Este é o caso deste trabalho, que relata a mineração de regras de associação em dados de uma pesquisa de interesse sobre cursos de especialização com alunos dos últimos períodos de cursos de graduação. O experimento teve como objetivo encontrar relações entre os cursos sob o ponto de vista do interesse dos alunos que estão perto de concluir a graduação, para indicar os cursos que a universidade oferecerá nos semestres seguintes. O uso de regras de associação nesse tipo de aplicação não é comum, e o presente trabalho também busca verificar a aplicabilidade dessa técnica ao problema.

A próxima seção deste artigo apresenta com mais detalhes a aplicação estudada. A seção 3 traz a mineração de regras propriamente dita e discute os resultados obtidos. Por fim, são apresentadas nossas conclusões e as prioridades para trabalhos futuros.

2. Aplicação: oferta de cursos *versus* interesse de alunos de graduação

A especialização é uma modalidade da pós-graduação *Lato Sensu* cuja oferta pelas Instituições de Ensino Superior (IES) tem crescido exponencialmente, especialmente nas instituições privadas (Gois, 2006). Em seus sítios na Internet, as IES oferecem diversos cursos de especialização, nas mais diversas áreas do conhecimento. Isso resulta em um grande número de possibilidades de cursos para cada profissional, o que aumenta a difi-

culdade de se prever a demanda potencial em cada projeto de curso. Assim, também é difícil, para as IES, selecionar os cursos a ofertar, já que uma decisão errada pode favorecer um curso pouco demandado em detrimento de outro, que poderia ter uma boa procura. Há também restrições de infra-estrutura e calendário, pois tais cursos disputam entre si não apenas a clientela, mas também horários e salas de aula. É preciso, portanto, desenvolver métodos que apóiem a tomada de decisão, evitando que a oferta de cursos seja feita simplesmente com base na intuição dos dirigentes das IES.

Para realizar o objetivo deste trabalho, foi feita uma pesquisa de interesse com os alunos dos dois últimos períodos dos cursos de graduação dessa unidade nas áreas escolhidas. Tais alunos são representativos dos interessados por cursos de pós-graduação já que, em nossa universidade, os recém-graduados representam 35% dos candidatos. Por já estarem na fase final de seus cursos, esses alunos em geral têm a capacidade de compreender o interesse do mercado de trabalho nas especialidades oferecidas na pós-graduação. As áreas escolhidas foram aquelas nas quais nossa universidade oferece o maior número de cursos de especialização: Informática (11 cursos), Direito (10 cursos) e Gestão (20 cursos). A pesquisa foi feita com 434 alunos, sendo 90 do curso de Sistemas de Informação, 203 de Direito, 67 de Administração de Empresas e 74 de Ciências Contábeis, sendo os dois últimos da área de gestão.

No formulário usado para coleta de dados, cada aluno deveria assinalar, dentre os cursos listados, aqueles em que teria interesse, não havendo limites mínimos ou máximos para a escolha. Também não foram feitas restrições quanto à marcação de cursos em áreas que não a do curso de graduação. Os dados foram organizados em tuplas com uma variável booleana por curso, correspondendo à marcação do curso pelo respondente. Cada tupla contém também o curso de graduação e o período cursado.

Apesar de a pesquisa ter mencionado todos os cursos de especialização que a universidade possui em cada área, a unidade onde os dados foram coletados não teria condições de abrigar todos eles, havendo a necessidade de se escolher apenas os de maior demanda potencial. Atualmente, há nessa unidade turmas em andamento dos cursos de Ciências Penais (Direito), Gestão Financeira (Gestão), Bancos de Dados (Informática) e Gerência de Projetos de Software (Informática), além de cursos de outras áreas de conhecimento não incluídas no levantamento.

A contagem das escolhas dos alunos ofereceu uma indicação da demanda potencial (Tabela 1). Os cursos de Direito Processual, Direito Tributário, Direito Público e Direito Constitucional são os mais procurados em sua área e, portanto, são os principais candidatos à oferta. No entanto, essa simples contagem não é suficiente para a decisão, pois os alunos poderiam marcar várias opções, e sem ordem de preferência (Tabela 2). Além disso, a decisão de oferta de um curso apenas por ser o mais indicado não garante um bom nível de inscrições, já que ele pode ser a segunda opção de vários alunos. Por exemplo, dos 79 alunos que marcaram Direito Constitucional, 38 também marcaram Direito Processual. Se o primeiro for oferecido, a procura pelo segundo pode cair para apenas 41 alunos, sobrando menos candidatos que outros três cursos.

3. Mineração de dados e análise de resultados

O objetivo da etapa de mineração de regras de associação entre os cursos foi apoiar a decisão de quais cursos ofertar, buscando esclarecer os dados de demanda potencial dos

cursos, como também reduzir o número de cursos, já que não haveria condições de oferecer todos. Através dessas regras se esperava determinar situações em que o interesse superposto poderia causar uma divisão da demanda potencial por cursos.

Tabela 1. Número de alunos que marcaram cada um dos cursos de Direito

Curso	Alunos
Ciências penais	50
Direito civil	51
Direito constitucional	79
Direito de empresa	33
Direito do trabalho	65
Direito eleitoral	21
Direito privado	19
Direito processual	101
Direito público	80
Direito tributário	93

Tabela 2. Marcação simultânea de dois dos cursos de Direito mais procurados

Cursos	Alunos
Direito Constitucional e Direito Processual	38
Direito Constitucional e Direito Público	34
Direito Constitucional e Direito Tributário	25
Direito Processual e Direito Público	34
Direito Processual e Direito Tributário	33
Direito Público e Direito Tributário	27

Para produzir elementos de suporte à decisão, optou-se por gerar regras de associação, usando o algoritmo Apriori (Agrawal et al., 1993). Apenas as regras de cardinalidade 2, em que a marcação de um determinado curso implica na marcação de outro foram consideradas. Considerando um suporte mínimo de 4% e uma confiança mínima de 40%, foram obtidas 20 regras de associação (Tabela 3). Esses índices foram estabelecidos por tentativa e erro, buscando fazer com que pelo menos uma regra de associação para os cursos de gestão aparecesse, já que nessa área as associações foram mais fracas. Em um futuro estudo, havendo uma maior quantidade de entrevistas, esses limiares devem ser revistos. O número de regras encontradas inclui as eventuais regras recíprocas do tipo $\{A\} \rightarrow \{B\}$ e $\{B\} \rightarrow \{A\}$.

É interessante observar a partir da Tabela 3 que Direito Processual é conseqüente dos cursos de Direito mais marcados, exceto Direito Tributário, porém não é antecedente, para esses níveis de suporte e confiança, de nenhum deles. Isso sugere que entre todos esses cursos de Direito os únicos que devem ser oferecidos são Direito Processual e Direito Tributário – ambos com um bom número de marcações e que podem atrair também alunos interessados nos outros principais programas. Na área de Informática, os principais conseqüentes são Gerência de Projetos de Software (53 marcações) e Banco de Dados (53 marcações), sendo, inclusive, recíprocos. Tal resultado, porém, pode ter sido influenciado pelo fato de que atualmente esses dois são os únicos cursos de Informática em andamento na unidade. A oferta dos outros dois cursos mais marcados, Gestão da Informação (49 marcações) e Segurança e Auditoria de Sistemas (36 marcações), parece arriscada, já que existem associações deles com Gerência de Projetos de Software e Banco de dados respectivamente. Na área de Gestão apareceram, para esses níveis de suporte e confiança, apenas os cursos Controladoria (39 marcações) e Gestão Tributária (44 marcações). O fato de um ser conseqüente do outro com níveis semelhantes de suporte e confiança não facilita a decisão de escolha entre eles, mas sugere que apenas um deles deve ser ofertado. Os outros dois cursos mais marcados dessa área, Gestão Financeira (50 marcações) e Gestão Estratégica de Negócios (41 marcações), também podem ser ofertados, já que não há risco representativo de concorrência entre eles.

Tabela 3. Regras de associação encontradas

Área	Antecedente	Conseqüente	Suporte	Confiança
Direito	Direito Constitucional	Direito Processual	8,8%	48%
	Direito Civil	Direito Processual	8,3%	71%
	Direito Constitucional	Direito Público	7,8%	43%
	Direito Público	Direito Constitucional	7,8%	43%
	Direito Público	Direito Processual	7,8%	43%
	Direito do Trabalho	Direito Processual	6,9%	46%
	Ciências Penais	Direito Processual	6,2%	54%
	Direito do Trabalho	Direito Tributário	6,2%	42%
	Direito Civil	Direito Constitucional	5,3%	45%
	Ciências Penais	Direito Constitucional	5,1%	44%
Informática	Banco de Dados	Gerência de Projetos de Software	6,5%	53%
	Gerência de Projetos de Software	Banco de Dados	6,5%	53%
	Redes de Computadores	Banco de Dados	5,5%	71%
	Banco de Dados	Redes de Computadores	5,5%	45%
	Gestão da Informação	Gerência de Projetos de Software	5,3%	47%
	Gerência de Projetos de Software	Gestão da Informação	5,3%	43%
	Engenharia de Software	Gerência de Projetos de Software	4,8%	66%
	Segurança e Auditoria de Sistemas	Banco de Dados	4,6%	56%
Gestão	Controladoria	Gestão Tributária	4,6%	51%
	Gestão Tributária	Controladoria	4,6%	46%

4. Conclusões

A mineração de regras de associação ajudou a produzir elementos para orientar a tomada de decisão sobre os cursos de especialização que a universidade deve ofertar. A detecção da associação potencial entre cursos indicou com clareza o risco de concorrência interna. Com isso, observa-se que nem todos os cursos que tiveram uma alta indicação de demanda devem ser ofertados, pois há o risco de que disputem os mesmos candidatos. É necessário determinar estratégias para a oferta dos cursos concorrentes em períodos distintos, ou mesmo em unidades distintas da universidade. Em um trabalho futuro, pretendemos fazê-lo usando *clustering*, com eliminação progressiva de cursos menos relevantes.

O estudo aqui apresentado tem caráter preliminar, e buscou obter elementos objetivos e úteis usando um instrumento de coleta de preenchimento rápido. Com isso, foi possível realizar um estudo rápido, voltado para a tomada de decisão em prazo muito curto, gerando decisões ágeis em um mercado rapidamente mutável e muito competitivo. Para essa finalidade, o uso de regras de associação mostrou ser adequado e suficiente. O estabelecimento de correlações entre as escolhas indicadas pelos alunos e outros parâmetros pessoais, tais como idade, emprego atual, renda, expectativa de empregabilidade e diversos outros, serão objeto de futuros esforços.

Referências

Agrawal, R.; Imielinksi, T.; Swami, A. Mining association rules between sets of items in large databases. Proceedings of the *1993 ACM SIGMOD Conference*, p.207-216. ACM Press, 1993.

Carvalho, L.A.V. *Data mining*: a mineração de dados no marketing, medicina, economia, engenharia e administração. Rio de Janeiro: Ciência Moderna, 2005.

Gois, A. *Setor privado já domina pós-graduação*. São Paulo: Folha de São Paulo, 31/07/2006. Disponível em: <http://www1.folha.uol.com.br/folha/educacao/ult305u18841.shtml>.