

Mineração de Dados na Formação de Turmas para a Recuperação Paralela na Educação Básica

Creusa Fernandes Correia¹ e Edson Pinheiro Pimentel¹

¹Universidade Federal do ABC (UFABC) – Santo André - SP
lenincorreia@ig.com.br, edson.pimentel@ufabc.edu.br

Abstract. *Parallel Recovery is one of the mechanisms that school uses to take into account the diversity of rhythms of the students' learning. This article aims to present the use of data mining techniques to form more homogeneous groups for the application of parallel recovery.*

Resumo. *A Recuperação Paralela é um dos mecanismos que a escola utiliza para atender à diversidade de ritmos de aprendizagem dos alunos. Este artigo tem por objetivo apresentar o uso de técnicas de mineração de dados para a formação de turmas mais homogêneas para a aplicação da recuperação paralela.*

1. Introdução

Nas últimas décadas a recuperação da aprendizagem dos alunos tem preocupado tanto os educadores que atuam na educação básica como as autoridades responsáveis pela implementação de políticas educacionais para esse nível de ensino [CENP, 2010]. Alunos com baixo rendimento tendem a ficar desmotivados e muitas vezes incrementam os números da evasão escolar.

No Estado de São Paulo, com a adoção do fim da reprovação entre séries e a opção por ciclos a partir de 1997, o projeto de implantação da progressão continuada deve especificar mecanismos que assegurem atividades de reforço e recuperação paralela contínuas, ao longo do processo [CEE/SP, 1997]. Atualmente, a cada avaliação periódica, o professor responsável deve registrar em ficha individual, as dificuldades observadas de aprendizagem bem como as recomendações aos próprios alunos, aos pais e outras providências a serem tomadas, ficando a seu critério indicar ou não os alunos com rendimento abaixo da média para estudos de recuperação.

Agrupar os alunos por série tem sido o critério utilizado na maioria das escolas para formação das turmas de recuperação. Nestes grupos encontram-se alunos com diversos Níveis de Aquisição de Conhecimentos (NAC), geralmente com lacunas distintas, sendo o mesmo conteúdo ministrado para todos. Acredita-se que a criação de grupos homogêneos de acordo com o NAC dos estudantes, permitiria um melhor planejamento para iniciar o resgate destes conteúdos para que cada aluno pudesse alcançar mais rapidamente o nível esperado para sua série de origem.

A identificação de grupos homogêneos a partir da análise humana tende a ser subjetiva caso não sejam identificados critérios claros para essa separação. Diversas áreas têm se utilizado de técnicas de mineração de dados para a identificação de grupos homogêneos através do reconhecimento de padrões [Witten e Frank, 2005].

Este artigo tem por objetivo apresentar o uso de técnicas de mineração de dados para a formação de turmas similares de recuperação paralela e está organizado como segue: A seção 2 descreve a técnica de agrupamento em Mineração de Dados utilizada nesse trabalho. A seção 3 apresenta o estudo de caso realizado. Por fim, a seção 5 apresenta as considerações finais e possíveis trabalhos futuros.

2. Mineração de Dados e Agrupamento

Mineração de dados é o processo de descoberta de conhecimento em bases de dados com a utilização de ferramentas computacionais a fim de descobrir informações valiosas, potencialmente úteis, descritas na forma de padrões, a partir dos volumes de dados coletados e armazenados pelas organizações [Fayyad, et. al., 1996].

Para direcionar a descoberta do conhecimento com qualidade e eficiência é preciso que os dados sejam previamente preparados envolvendo etapas de pré-processamento como limpeza, integração, transformação e redução de dados [Witten e Frank, 2005]. Detalha-se a seguir a tarefa de Agrupamento (clusterização) utilizada no estudo de caso que será descrito na seção 3. Os resultados apresentados nos trabalhos de Pimentel et al (2003) confirmam que técnicas de clusterização são bastante úteis para a formação de grupos homogêneos de aprendizes.

A análise de cluster é um procedimento multivariado para detectar grupos nos dados. Através da clusterização dos dados, pode-se classificar um conjunto de objetos dentro de grupos ou categorias. Dentre as técnicas de clusterização destaca-se o método K-means que usa o algoritmo de agrupamento de dados por k-médias cujo objetivo é encontrar a melhor divisão de P dados em K grupos C_i , $i = 1, \dots, K$, de maneira que a distância total entre os dados de um grupo e o seu respectivo centro, somada por todos os grupos, seja minimizada [Witten e Frank, 2005].

3. Estudo de Caso

Para o estudo de caso foram selecionados os dados referentes ao desempenho dos alunos com rendimento abaixo da média de duas unidades escolares de 1º e 2º anos do Ensino Médio, perfazendo um total de 250 registros. O Sistema de Registro de Rendimento Escolar utilizado nestas escolas é um sistema *on-line* no qual o professor registra o rendimento e a frequência para cada estudante. Ao digitar uma nota abaixo da média o identificador do aluno aparece em destaque para que o professor acione a tela de Desempenho a fim de indicar as principais dificuldades do aluno, as recomendações do professor ao aluno, aos pais e as providências para auxiliar o aluno.

O ambiente utilizado neste estudo de caso para aplicação das técnicas de mineração de dados foi o WEKA (Waikato Environment for Knowledge Analysis), desenvolvido pela Universidade de Waikato da Nova Zelândia [University of Waikato, 2004]. A seguir descreve-se a aplicação das etapas do processo de descoberta de conhecimento

4.1. Preparação dos Dados

Após seleção e limpeza os dados foram transformados para se enquadrar ao formato recomendado para o WEKA. A tabela 1 apresenta os atributos selecionados para este estudo de caso e a sua descrição.

Tabela 1. Atributos coletados no modelo relacional

Atributo	Tipo de dado	Descrição
Inter	Nominal (sim, não)	Dificuldade de interpretação
Rac	Nominal (sim, não)	Dificuldade de raciocínio lógico
Conc	Nominal (sim, não)	Não produz textos com concordância
Argu	Nominal (sim, não)	Dificuldades de argumentação
Compr	Nominal (sim, não)	Falta de compromisso com os estudos
Recup	Nominal (sim, não)	Recomendado para a recuperação paralela

4.2. Extração de Padrões e Análise de Resultados

Durante o treinamento foram feitos agrupamentos de 2 até 10 clusters (grupos). Identificou-se que o agrupamento com 5 clusters resultou em centróides mais determinantes para formação de turmas homogêneas.

Os centróides (centro do cluster) são pontos escolhidos aleatoriamente de acordo com o número k de agrupamento especificado pelo usuário. Depois da definição dos centróides é analisado cada elemento do conjunto e a partir da medida de distância entre o cluster e o dado elemento é definido a qual cluster ele pertencerá. O elemento vai ser inserido no cluster do qual tiver menor distância. Na próxima etapa, é calculado um novo centróide para cada cluster, pois os clusters escolhidos inicialmente não são os centros definitivos. Essas tarefas se repetem até que os centróides não sejam mais alterados [Witten e Frank, 2005].

A figura 1 apresenta os centros de cada grupo sendo cada um deles interpretados a seguir:

Attribute	Full Data (250)	Cluster#				
		0 (48)	1 (88)	2 (16)	3 (78)	4 (20)
INTER	sim	não	sim	sim	sim	não
RAC	não	não	sim	não	não	sim
CONC	não	não	não	não	sim	não
ARGU	não	não	não	não	sim	não
COMPR	sim	sim	não	sim	sim	não
RECUP	sim	não	sim	sim	sim	sim

Figura 1. Agrupamento em clusters.

- Cluster 0 - Neste cluster com 48 alunos (19,2%) nota-se que apesar do rendimento abaixo da média apresentaram apenas falta de compromisso com as atividades e avaliações, não tendo sido recomendados para a recuperação.
- Cluster 1 - Foram agrupados nesse cluster 88 alunos (35,2%) com problemas de interpretação e raciocínio lógico, indicados para recuperação caracterizando assim uma turma de recuperação em matemática.
- Cluster 2 - Representa aqueles alunos com dificuldades apenas de interpretação, podendo indicar dificuldades de interpretação de textos, de problemas, formando uma turma interdisciplinar, com conteúdos específicos de interpretação tanto de português como de matemática. Nele constam 16 alunos (6,4%).

- Cluster 3 - Este cluster forma uma turma de recuperação em português, apresentando 78 alunos (31,2%) com dificuldades de interpretação, concordância argumentação e falta de compromisso.
- Cluster 4 - Assim como no cluster 1, este forma uma turma de recuperação em matemática, que difere apenas com alunos que não apresentam problemas de interpretação, podendo ser trabalhado atividades que desenvolvam o raciocínio lógico. Nele constam 20 alunos (8%).

4. Conclusão e Trabalhos futuros

A formação de turmas para recuperação paralela na Educação Básica tem sido motivo de muitas reuniões pedagógicas nas unidades escolares. Agrupamentos baseados apenas nas indicações de dificuldades de aprendizagem apresentadas pelos professores não tem se mostrado adequado.

Neste trabalho a técnica de clusterização conseguiu agrupar os alunos pelos problemas de aprendizagem apresentados, inclusive, separando em um grupo específico aqueles alunos com dificuldades de interpretação, podendo assim formar uma turma interdisciplinar. Não é possível concluir ainda se os resultados obtidos com a técnica de classificação são significativos. No entanto, esses resultados podem ser utilizados para apoiar as reuniões pedagógicas e a tomada de decisão na formação de grupos mais homogêneos.

Como trabalhos futuros pretende-se integrar módulos de avaliação a ferramentas de mineração de dados a fim de não apenas indicar os alunos com dificuldades de aprendizagem para recuperação paralela mas fazê-lo com dados mais significativos.

5. Referências

- CENP (2010) Equipe Técnica – EF Ciclo II e EM – Processos de Recuperação: Avanços e desafios na nova Resolução. Planejamento Escolar.
- CEE/SP (1997) Conselho Estadual de Educação de São Paulo. Deliberação nº 09. Institui no Sistema de Ensino do Estado de São Paulo o regime de progressão continuada no ensino fundamental. Disponível em http://www.ceesp.sp.gov.br/Deliberacoes/de_09_97.htm. Acesso em 25/7/2011.
- FAYYAD, Usama M.; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. (1996) Data Mining to knowledge discovery: an overview. *Advances in Knowledge Discovery and Data Mining*.
- PIMENTEL, Edson P.; FRANÇA, Vilma F.; OMAR, Nizam. (2003) A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. XIV Simpósio Brasileiro de Informática na Educação.
- University of Waikato. (2004) Weka 3 – Machine Learning Software in Java. Disponível em <http://www.cs.waikato.ac.nz/ml/weka>. Acesso em 1/8/2011.
- WITTEN, Ian H. e FRANK, Eibe. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, 2 edition.