

ESTUDO DO DESENVOLVIMENTO COGNITIVO INDIVIDUAL E DE GRUPOS ATRAVÉS DA ANÁLISE AUTOMÁTICA

Maria de Fátima W. do Prado Lima – CCTI-UCS - mfwplima@ucs.br Carine G. Webber – CCTI-UCS - cgwebber@ucs.br Bruno B. Guimarães - CCTI-UCS - bbguimar@ucs.br

Resumo. Este artigo propõe o uso de métodos automáticos de análise de dados para possibilitar a integração das teorias de ensino-aprendizagem aos ambientes virtuais de aprendizagem. Em domínios onde o desenvolvimento cognitivo passa pela resolução de problemas, a análise do processo de aprendizagem depende do raciocínio desenvolvido pelo aluno para chegar à solução e do devido acompanhamento do professor. O acompanhamento deste processo permite reconhecer padrões cognitivos que podem desencadear o sucesso ou o fracasso na aprendizagem, auxiliando o professor nas suas tarefas. Com esta finalidade, este artigo propõe o uso de algoritmos de clustering para auxiliar a análise do desenvolvimento cognitivo individual e de grupos através de um experimento na aprendizagem da lógica de programação.

Palavras-chave: Tecnologias na Educação, Trabalho em grupo, Colaboração.

A STUDY ON INDIVIDUAL AND COLLECTIVE COGNITIVE DEVELOPMENT THROUGH AUTOMATIC ANALYSIS

Abstract. This article proposes the use of automatic methods for data analysis in order to integrate educational theories to virtual learning environments. In domains where the cognitive development relies on problem solving, the analysis of the learning process depends on the reasoning developed by learners to get the solution and on the monitoring of the teacher. The accompaniment of this process allows recognizing cognitive standards that can trigger off the success or the failure of learning, assisting the teacher in this task. This article proposes the use of clustering algorithms to the analysis of individual and groups cognitive development through an experiment in programming logic learning.

Keywords: *Educational Technology, Group work, Collaboration.*

1. Introdução

O ensino da área de programação de computadores exige que o professor consiga analisar o desenvolvimento de raciocínio dos alunos a fim de auxiliá-los no processo de



ensino-aprendizagem. Apesar da utilização de ambientes virtuais de aprendizagem estar em evidência desde a última década, várias premissas utilizadas como base no processo de ensino-aprendizagem precedem este período. Foi Piaget (2000) que em sua teoria do desenvolvimento cognitivo evidenciou a abstração lógica e matemática como sendo um processo reflexivo, gerado por uma reflexão interna e de natureza metacognitiva. Antes dele Vygotsky (1962) afirmava que o desenvolvimento deve ocorrer explorando as funções cognitivas em vias de maturação, ao invés de explorar as funções já amadurecidas. Mas este processo depende fundamentalmente da interação social, onde um indivíduo mais capaz auxilia o indivíduo em fase de desenvolvimento. De forma complementar, Gardner (1989) enfatiza que para realizar a avaliação cognitiva do crescimento do aluno não basta classificar, aprovar ou reprovar os alunos. Deve-se conhecer o processo de aprendizado, informando ao aluno sobre a sua capacidade e o professor tendo a certeza do que foi aprendido. Apesar das teorias citadas serem distintas, e algumas vezes consideradas controversas, elas refletem características que a utilização de um ambiente virtual de aprendizagem na área de programação apresenta. Primeiro, o aluno ao desenvolver um algoritmo desencadeia um raciocínio lógico complexo. Segundo, para poder auxiliar no desenvolvimento do aluno, o professor deve acompanhar todo o processo de lógica realizada. Terceiro, a avaliação não deve considerar apenas o resultado final, mas todo o processo de aprendizagem. Além disso, deve ocorrer a interação entre professor e aluno e entre os próprios pares para aumentar a efetividade do processo de ensino-aprendizagem. Ou seja, o ambiente virtual deve fornecer ao professor os indicadores necessários para que ele interfira de forma correta no processo de aprendizagem do aluno e, ao mesmo tempo, agilizar a interação entre os pares, de forma com que os alunos sejam levados a interagir com os pares que mais podem contribuir com o seu estágio atual de desenvolvimento.

A fim de fornecer tantos subsídios complexos, o ambiente virtual de aprendizagem deve ser dotado de certo grau de inteligência, agilizando principalmente a interação entre os pares corretos, formando os grupos de trabalho de acordo com o estágio de desenvolvimento cognitivo dos indivíduos. Se a análise de tais indicadores for realizada de forma manual e individual, uma visão detalhada dos fenômenos pode ser constatada, mas este processo irá demandar um tempo considerável e naturalmente inaceitável em termos das ações que o professor deve realizar no processo de ensino-aprendizagem. Para analisar um grande número de critérios, obter uma boa visão global das informações, em um tempo hábil, os métodos automáticos constituem um recurso precioso capaz de sintetizar os resultados. Neste cenário, técnicas e ferramentas automatizadas auxiliam na transformação de dados em informações significativas e em conhecimento para o educador. Os resultados produzidos por análises automáticas permitem que ocorra o acompanhamento do desenvolvimento cognitivo, tanto em nível individual quanto de grupos.

2. Pesquisas Relacionadas

Aplicações dos métodos de análise de dados na Educação visando o acompanhamento de processos de aprendizagem ainda são recentes (Castro et.al, 2007). Em geral as pesquisas por técnicas e ferramentas automatizadas têm sido realizadas no intuito de formar grupos nos ambientes virtuais a partir das características individuais, do conhecimento do sujeito e de seu comportamento. Algoritmos de *clustering* constituem as técnicas mais aplicadas para a formação automática de grupos (Pollalis e Mavrommatis, 2009; Anaya e Boticario, 2009; Hogo, 2010). Os trabalhos de Maite (2008) e Christodoulopoulos (2007) demonstram como técnicas de mineração contribuem para a formação e análise de grupos. Maite utiliza tais técnicas para auxiliar



no diagnóstico do aluno em um ambiente educacional *web* denominado Magadi. Em outra direção, Christodoulopoulos aborda a importância da formação de grupos homogêneos e heterogêneos dentro de ambientes de aprendizados colaborativos e adaptativos, realizando alguns testes utilizando algoritmos de *clustering* para formação de grupos.

Perera (2009) descreve a utilização de técnicas de *clustering* para monitorar as ações realizadas pelos grupos dentro de um curso de desenvolvimento de software, onde foram consideradas principalmente as contribuições realizadas no *wiki* e as ações registradas em um sistema de gerenciamento de tarefas. Anaya (2009) utilizou técnicas automáticas para analisar o nível de colaboração dos estudantes nos grupos e individualmente, avaliando variáveis tais como: número de conversações iniciadas pelo aluno, número de mensagens enviadas e número de mensagens respondidas pelo usuário. Ambos os trabalhos utilizaram o algoritmo EM através da ferramenta WEKA (Witten e Frank, 2005).

3. Técnicas de Clustering

A tarefa da mineração de dados consiste em um processo de extração automática de padrões válidos, novos, potencialmente úteis e compreensíveis, visando melhorar o entendimento de um problema ou apoiar uma tomada de decisão (Fayyad, 1995; Mitchel, 1997; Castro el.al, 2007). Esse processo envolve a preparação de dados, a aplicação de algoritmos de mineração de dados, a análise e interpretação de padrões, e a avaliação do conhecimento extraído. Em um conjunto de instâncias não classificadas, as técnicas de *clustering* realizam a partição dos dados em grupos de instâncias (*clusters*) que apresentam grandes similaridades entre si e pouca similaridade com instâncias de outros *clusters*. Os *clusters* formados devem ser interpretados em termos de suas características mais relevantes, normalmente apresentadas pelo elemento centroide (ponto médio) do *cluster*.

No contexto deste trabalho, o acompanhamento da aprendizagem através de algoritmos de *clustering* pode ser útil fornecendo informações para: formação de grupos homogêneos, identificação de características representativas dos grupos, acompanhamento temporal da aprendizagem, identificação de indivíduos que fogem aos padrões do grupo, entre outras.

O ponto de partida do uso de técnicas de *clustering* é a disponibilidade de dados históricos que representem cenários do domínio a ser estudado. De modo geral, algoritmos de *clustering* geram partições dos dados a partir de similaridades entre valores de atributos. A visualização gráfica das partições é um recurso necessário e desejável para a análise dos resultados obtidos. Uma vez geradas, as partições de dados devem ser interpretadas no contexto do domínio e favorecer *insights* e reflexões que conduzam o usuário à validação de hipóteses e construção de conhecimento.

O método particional mais conhecido e empregado na formação de *clusters* é o *k-means*. O algoritmo *k-means* utiliza o conceito estatístico do centroide para determinar a semelhança das instâncias de um *cluster*. Em muitos casos o algoritmo *k-means* é combinado a outras técnicas que refinam e melhoram a formação dos grupos. O algoritmo Expectativa (EM), proposto por Dempster, Laird e Rubin (1977), é uma destas técnicas. Em estatística, o EM funciona como um método para identificar a máxima semelhança entre parâmetros de modelos estatísticos, sendo que variáveis do modelo podem não estar acessíveis a observação. O algoritmo EM é reconhecido por produzir bons resultados de formação de *clusters*, sendo portanto utilizado nesta trabalho. A partir do estudo sobre algoritmos de *clustering* desenvolveu-se um



experimento que teve por objetivo analisar o seu uso no acompanhamento dos processos de ensino e aprendizagem de uma disciplina introdutória de programação de computadores.

4. Método de Pesquisa

A fim de avaliar o uso de métodos automáticos para avaliação dos processos de aprendizagem, preparou-se um experimento com alunos da disciplina de Introdução à Programação (84 horas) da Universidade de Caxias do Sul. O objetivo da disciplina é preparar o aluno para que ele seja capaz de analisar problemas e formular soluções algorítmicas estruturadas que possam ser executadas pelo computador. A turma composta por 20 alunos participou do experimento desenvolvido em três etapas. A primeira etapa aconteceu após 36 horas de curso regular, a segunda após 40 horas e a terceira após 44 horas. Optou-se por realizar as três etapas do experimento com pouco intervalo de tempo entre elas para avaliar a evolução dos alunos em tarefas variadas e verificar a sensibilidade dos métodos de análise automática. O nível de complexidade dos problemas trabalhados foi o mesmo nas três etapas. A tabela 1 apresenta a descrição das etapas do experimento.

Etapa do experimentoTipo de tarefaConfiguração1Resolução de problemaIndividual2Análise de soluçãoDupla3Análise de soluçãoIndividual

Tabela 1. Descrição das etapas do experimento

Na etapa 1 solicitou-se aos alunos a resolução de um problema de média complexidade envolvendo estruturas condicionais e iterativas. Nesta etapa avaliou-se as habilidades de interpretação, percepção e raciocínio lógico. As soluções dos alunos foram analisadas e produziu-se uma biblioteca de erros (19 erros). Nesta primeira análise os erros foram identificados manualmente, porém a sua descrição permitirá que posteriormente seja desenvolvido um sistema de reconhecimento automático. As soluções dos alunos foram descritas segundo os erros identificados de forma a compor um *dataset*. Um *dataset* corresponde a um arquivo binário onde cada linha representa um caso ou instância. A tabela 2 ilustra um exemplo de *dataset* onde cada linha representa os erros encontrados na solução de cada aluno: 0 indica erro ausente, 1 indica erro presente.

Tabela 2. Modelo de dataset desenvolvido na etapa 1

	erro1	erro2	erro3	erro4	erro5	erro6	erro7	erro8	erro9	erro10
aluno1	1	0	0	1	1	0	0	1	1	0
aluno2	0	1	1	1	1	0	0	0	0	0
aluno3	1	0	0	1	1	0	0	0	1	0

O *dataset* da primeira etapa foi submetido ao algoritmo EM, disponível para uso na ferramenta WEKA (Witten et.al., 2011). A ferramenta WEKA é um software livre,



que implementa e disponibiliza o algoritmo EM. O algoritmo foi capaz de reconhecer 2 agrupamentos que corresponderam a 2 padrões de desenvolvimento cognitivo dos alunos. O primeiro *cluster*, contendo cerca de 80% dos alunos, foi reconhecido como o grupo de alunos com conhecimentos básicos de lógica, porém com dificuldades de compreensão do enunciado (nível fraco). No segundo *cluster* foram agrupados alunos com maiores dificuldades e muitos erros de lógica (nível muito fraco).

Na segunda etapa do experimento, os alunos foram agrupados em duplas. Cada dupla foi formada contendo um aluno de cada *cluster* anteriormente formado, buscando a complementaridade dos conhecimentos demonstrados. Na segunda etapa o desenvolvimento cognitivo de cada aluno foi novamente analisado e gerou um novo *dataset*. Na terceira etapa, os alunos realizaram a última atividade individual, o que permitiu a geração do terceiro *dataset*. A partir da geração dos *datasets* iniciou-se o trabalho de análise automática do desenvolvimento cognitivo individual e coletivamente dos alunos. Cada *dataset* foi submetido ao algoritmo EM, gerando *clusters* e análises apresentados na seção seguinte.

5. Resultados e Discussão

A fim de acompanhar a aprendizagem dos alunos de forma individual e também coletiva, utilizou-se uma escala para indicar o nível de conhecimento dos alunos sendo: (1) muito fraco, (2) fraco, (3) intermediário e (4) forte. Os alunos foram qualificados de acordo com o agrupamento a que pertenceram em cada etapa do experimento. A fim de analisar os resultados de forma coletiva desenvolveu-se também um *dataset* contendo os dados individuais de todos os alunos conforme tabela 3.

Etapa 1Etapa 2Etapa 3aluno1fracointermediáriofortealuno2muito fracointermediáriointermediárioaluno3fracofracointermediário

Tabela 3. Modelo de *dataset* desenvolvido no experimento

O dataset foi submetido ao algoritmo EM. A análise automática formou 5 clusters ou 5 agrupamentos distintos de alunos. Cada agrupamento foi associado a um padrão de comportamento de sucesso que revela aprendizagem ou de insucesso, o que revela dificuldades na aprendizagem. A tabela 4 sintetiza os padrões de aprendizagem identificados em cada *cluster* formado pela análise automática.



Cluster	Nº de alunos	Interpretação dos Clusters	
1	5	Alunos com facilidade de aprendizagem	
2	1	Alunos com grande dificuldade de raciocínio lógico não demonstrando evolução	
3	8	Aluno com ritmo lento de aprendizagem mas apresentano evolução	
4	4	Alunos com grande dificuldade inicial de aprendizagem porém demonstrando alguma evolução	
5	2	Dificuldade inicial porém grande capacidade de aprendizagem	

Tabela 4. Interpretação dos grupos formados

Observando de forma detalhada o *cluster* 1 nota-se que os alunos obtiveram um avanço no nível de conhecimento (eixo vertical) a partir da segunda etapa do experimento (figura 1, à esquerda). Mesmo tendo partido de níveis fraco (2) ou muito fraco (1), todos evoluíram. À direita da mesma figura 1 pode-se perceber a análise individual do aluno 5, demonstrando os níveis de conhecimento identificados a cada etapa do experimento. A análise dos dados a partir da formação de *clusters* permite que se observe comportamentos típicos, que atestam que o processo de aprendizagem do grupo e do indivíduo. Em casos especiais, a análise automática é capaz de identificar comportamentos que fogem dos demais padrões detectados. Um exemplo disto pode ser visto no segundo *cluster* formado, o qual envolveu apenas 1 aluno (figura 2).

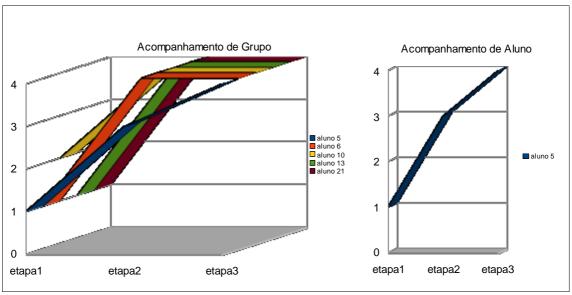


Figura 1 – Acompanhamento do cluster 1 (coletivo) e do aluno 5 (individual)

O aluno (23) foi o único no experimento que não demonstrou nenhuma evolução. O aluno teve muitas dificuldades na etapa 1 do experimento, apresentou uma avaliação melhor na atividade em dupla (possivelmente relacionado ao conhecimento complementar do outro aluno), não apresentando evolução na terceira etapa do experimento. A análise automática foi correta não incluindo este aluno em nenhum outro cluster.



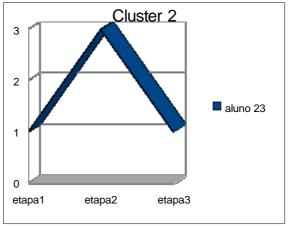


Figura 2 – Acompanhamento do cluster 2

O *cluster* 3 agrupou 8 alunos que iniciaram o experimento no nível fraco (2) e demonstraram uma evolução a partir do experimento colaborativo (etapa 2) conforme pode ser observado na figura 3. Mesmo se atividade colaborativa da etapa 2 não tenha influenciado o resultado no desenvolvimento cognitivo da maioria dos alunos, a partir dela houve uma evolução significativa (6 alunos). Somente os alunos 16 e 17 não demonstraram nenhuma evolução na etapa 3.

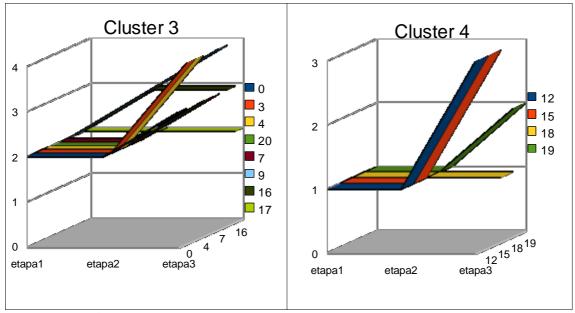


Figura 3 – Padrões de aprendizagem dados pelos clusters 3 e 4

O cluster 4 agrupou quatro alunos que iniciaram no nível muito fraco (1) de conhecimento (figura 3). Neste cluster os alunos agrupados representam um padrão de aprendizagem lento, de pouca evolução, atingindo no máximo o nível intermediário de conhecimento. Somente um aluno (18) não apresentou nenhuma evolução após a atividade colaborativa da etapa 2. O cruzamento dos dados dos 5 clusters permite gerar a visualização completa da turma de alunos (figura 4). Pode-se observar que com exceção do cluster 1, todos os demais representam a evolução dos alunos. Percebe-se também que a atividade colaborativa foi teve uma contribuição significativa para a



melhoria no desenvolvimento cognitivo dos alunos. Após esta atividade, 4 dos 5 grupos apresentaram uma evolução média maior que nas etapas anteriores.

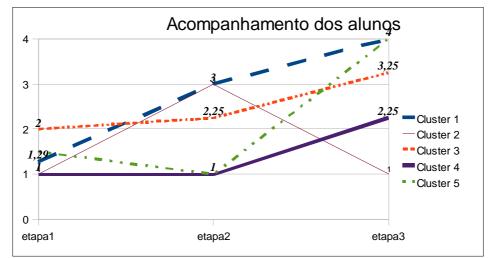


Figura 4 – Acompanhamento dos alunos através de valores médios dos clusters

Considera-se que neste experimento o algoritmo EM reconheceu com alto grau de sensibilidade perfis de alunos, cuja visualização é facilitada pela projeção gráfica.

6. Conclusão

Este artigo propôs um estudo sobre a aplicação de métodos automáticos de análise de dados que permitem que o professor acompanhe o processo de aprendizagem de cada aluno. O experimento desenvolvido permitiu também reconhecer grupos de alunos com o mesmo estilo de aprendizagem e portanto com padrões de comportamento similares. Em termos computacionais, o algoritmo EM se revelou um método de identificação de padrões de comportamento, apresentando resultados satisfatórios. Consideramos que ele pode ser potencialmente aplicado em cenários onde o professor planeje atividades em grupos (homogêneos ou heterogêneos). A identificação automática dos perfis de comportamentos dos alunos permite que o professor possa tomar decisões ágeis de acordo com o desenvolvimento cognitivo atual do aluno a cada etapa da aprendizagem.

Através da análise de gráficos e dados obtidos neste experimento foi possível reconhecer a potencialidade das técnicas de *clustering*. O acompanhamento em nível individual permite constatar o sucesso do processo de aprendizagem dado as dificuldades iniciais do aluno. Considerando os grupos, é possível extrair comportamentos típicos nas etapas de aprendizagem e assim construir-se perfis de alunos.

Referências

ANAYA, A.,BOTICARIO, J.G. (2009). Clustering learners according to their collaboration. In Proceedings of the 13th International Conference on Computer Supported Cooperative Work in Design, IEEE Press, Los Alamitos, pp. 540-545.

CASTRO, F., VELLIDO, A., NEBOT, A., MUGICA, M. (2007) Applying Data Mining Techniques to e-Learning Problems, Studies in Computational Intelligence (SCI) 62, p.183–221.



CHANCHARY, F., HAQUE, I., KHALID, S. (2008). Web Usage Mining to Evaluate the Transfer of Learning in a web-Based Learning Environment. In Proceedings of the First International Workshop on Knowledge Discovery and Data Mining, 249-253.

CHRISTODOULOPOULOS, C.E., PAPANIKOLAOU, K.A.A. (2007). Group Formation Tool in an E-Learning Context, Tools with Artificial Intelligence, In Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence, 2, pp.117-123.

DEMPSTER, A.P., LAIRD, N.M. & RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B*, pp.1–38.

FARIA, E.S.J., ADAN-COELLO, J. M., YAMANAKA, K. (2006). Forming groups for collaborative learning in introductory computer programming courses based on students' programming styles. In *Proceedings of the 36th ASEE/IEEE Frontiers in Education Conference*, San Diego, CA, pp. 6-11.

FAYYAD, U. M., PIATESKY-SHAPIRO, G. & SMYTH, P. (1995). From Data Mining to Knowledge Discovery: An Overview, em Advances in Knowledge Discovery and Data Mining. AAAI Press.

GARDNER, H., HATCB, T. Multiple intelligences go to school: educational implications of the theory of Multiple Intelligences. Educational Researcher, v.18, n.8. p.4-10, 1989.

HOGO, M.A. (2010). Evaluation of E-Learners Behaviour using Different Fuzzy Clustering Models: A Comparative Study. *International Journal of Computer Science and Information Security*, 7(2), pp.131-140.

ISOTANI, S., INABA, A., IkEDA, M., MIZOGUCHI, R. (2009). An ontology engineering approach to the realization of theory-driven group formation. *International Journal of Computer-Supported Collaborative Learning*, 4(4), pp.445-478.

MITCHELL, T. (1997). Machine Learning. McGraw Hill, 1997.

PERERA, D., KAY, J., KOPRINSKA, I., YACEF, K., ZAIANE, O. (2009). Clustering and Sequential Pattern Mining of Online Collaborative Learning Data. In *IEEE Transactions on Knowledge and Data Engineering*, 21(6), pp. 759-772.

PIAGET, J. (1995). Sociological Studies (L. Smith; L. Smith; et al. Trans.) London: Routledge.

PIAGET, J. Biologia e Conhecimento. Petrópolis: Editora Vozes, 2000.

POLLALIS,Y., MAVROMMATIS, G. (2009). Using similarity measures for collaborating groups formation: A model for distance learning environments. *European Journal of Operational Research*, 193, pp. 626-636.

VYGOTSKY, L. V. (1980) Mind in Society: The Development of Higher Psychological Processes. Harvard Univ Press.

VYGOTSKY, L.S. Thought and Language. Cambridge, MA: The MIT Press, 1962

WITTEN, I. H., FRANK, E., HALL, M.A. (2011) Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edtion. San Francisco: Morgan Kaufmann, 2011. 629p.