

Identificação de Grupos de Alunos em Ambiente Virtual de Aprendizagem: Uma Estratégia de Análise de Log Baseada em Clusterização

Márcia F. Pinheiro¹, Luiz Cortinhas F. Neto¹, Haroldo Nazaré de Sá Junior¹
Eulália C. da Mata¹, Antonio F. L. Jacob Jr.¹, Ádamo de Lima Santana¹

¹Instituto de Tecnologia – Universidade Federal do Pará (UFPA)
Caixa Postal 479 – 66075-110 – Belém –PA – Brasil

{eng.marciafontes, luizcf14, hnsj86, eucmata}@gmail.com,
{jacobjr, adamo}@ufpa.br

Abstract. *Have information about the learning process is of utmost importance for educators and students, as it allows to support decision making and reflection regarding teaching methodologies and contents used and performance of students. In this study were used Mining techniques in educational data in order to find groups of students from a large-scale educational program by analyzing a Distance Learning platform logs. The results obtained allow making qualitative and quantitative analysis of the use of the Moodle platform to groups of students found.*

Resumo. *Ter informações a respeito do processo de aprendizagem é de extrema importância para educadores e alunos, pois permite apoiar a tomada de decisão e reflexão quanto às metodologias aplicadas no ensino, bem como conteúdos utilizados e desempenho de alunos. Neste trabalho, foram utilizadas técnicas de mineração de dados educacionais com o objetivo de encontrar grupos de alunos de um programa educacional de larga escala através da análise de log de uma plataforma de Educação a Distância como parte de pesquisa motivada pela verificação de aproveitamento dos alunos na formação. Os resultados obtidos permitem fazer análise qualitativa e quantitativa do uso da plataforma pelos dos grupos de alunos encontrados.*

1. Introdução

Com o advento das Tecnologias da Informação e Comunicação (TICs), Internet, dos sistemas de informação para Web e dos recursos multimídia, surgiram diversas ferramentas e plataformas com a proposta de Web 2.0.

No campo da educação, a utilização desses sistemas de ensino eletrônico – e-learning – proporcionam novas possibilidades de aprendizagem na metodologia de ensino. Esses sistemas são denominados como Ambientes Virtuais de Aprendizagem (AVAs) por serem plataformas que centralizam ferramentas, mas que permitem interação entre os usuários de forma assíncrona e síncrona através de Wiki, chat, fórum, etc.

Os AVAs armazenam todas as interações dos usuários dentro da plataforma, informações do desempenho do aluno, podendo guardar informações como quais atividades um estudante participou, materiais que foram lidos e escritos, testes ao qual foi submetido, chats que o mesmo participou, páginas acessadas na plataforma, etc.

[Mostow et al., 2005], assim como informações pessoais sobre usuário, tais como perfil, em sua base de dados.

Dentre os AVAs, temos o Moodle, que é um sistema *open source* desenvolvido pelo australiano Martin Dougiamas para uso acadêmico. Engloba diversas ferramentas web 2.0 para facilitar a interação entre os usuários [Mata et al., 2010] e é a plataforma mais utilizada no mundo em educação a distância [CAPTERRA, 2014].

De maneira geral, os AVAs armazenam uma grande quantidade de informação o qual é muito valiosa para analisar o comportamento dos estudantes [Mostow and Beck, 2006] e que pode ser obtida através de Mineração de Dados, também conhecida como Descoberta do Conhecimento em Base de Dados (DCBD), da expressão em inglês Knowledge Discovery in Databases (KDD), que é a extração automática de padrões implícitos e interessantes a priori desconhecidas e potencialmente úteis a partir de grandes volumes de dados [Klösgen and Zytkow, 2002] [WITTEN, FRANK and HALL, 2005]. Para encontrar padrões nesses dados são envolvidos a Mineração de Dados (*Data Mining*), extração do conhecimento, descoberta da informação e padrão de processamento dos dados [Fayyad, 1996]. Baker e de Carvalho (2010) complementam que a Mineração de Dados permite descobrir novas informações no processo de identificação de relações entre dados que podem produzir novos conhecimentos e gerar novas descobertas científicas. Informações estas que podem ser muito úteis para tomada de decisão.

KDD tem sido aplicada em diversas áreas do conhecimento, como por exemplo, em finanças [KOVALERCHUK and VITYAEV, 2005], bioinformática [HU, 2011], combate ao crime e terrorismo [OKONKWO AND ENEM, 2011], saúde [GOSAIN and KUMAR, 2009], esportes [WICKER AND BREUER, 2010], etc. A Informática na Educação é uma linha de pesquisa que tem sido consolidada, como apresenta o trabalho de Frascareli e Pimentel (2012), e tem sido um tema estudado por diversos pesquisadores da área, em particular da Inteligência Artificial Aplicada à Educação [BARKER, ISOTANI and De Carvalho, 2011].

KDD tem sido utilizada com o intuito de investigar perguntas científicas na área de educação como, por exemplo, quais são os fatores que afetam a aprendizagem? Como desenvolver sistemas educacionais mais eficazes? Ou ainda a relação da abordagem pedagógica e o aprendizado do aluno, estas informações podem ser úteis não somente para os educadores, mas também aos próprios alunos, uma vez que pode ser orientada para diferentes fins por diferentes participantes no processo. Dentro deste contexto surgiu a Mineração de Dados Educacionais (do inglês, Educational Data Mining - EDM), que é definida como a área de pesquisa que tem como principal foco o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais, desta maneira é possível compreender de maneira mais eficaz e adequada os alunos, como estes aprendem, o papel do contexto na qual a aprendizagem ocorre, além de outros fatores que influenciam a aprendizagem.

Este trabalho propõe como metodologia à aprendizagem de log (JANSEN, 2006) adaptada para plataforma Moodle utilizada em grande escala para cursos de incentivos públicos e privados semi-presenciais brasileiros, aplicando clusterização através da abordagem de log para encontrar grupos de alunos na plataforma bem como o caminho médio de cada grupo na plataforma, como trabalho inicial de verificação de aproveitamento dos alunos na formação.

Na Seção 2 são apresentados trabalhos que utilizaram técnicas de EDM; a Seção 3 apresenta a contextualização do domínio o qual esse trabalho realizou a pesquisa; Na Seção 4 são apresentados os experimentos realizados; Na Seção 5 são apresentados os resultados obtidos e a discussão dos mesmos e na Seção 6 são feitas as conclusões e possíveis trabalhos futuros.

2. Trabalhos Relacionados

Como área de pesquisa em crescimento [Frascareliand Pimentel,2012], diversas pesquisas têm sido feitas no campo da EDM utilizando dados de AVAs. Por exemplo, Gottardo, Kaestner e Noronha(2012) aplicaram EDM utilizando classificação em busca da obtenção de modelos para inferir e prever o desempenho dos estudantes em um curso de Educação a Distância (EAD) a partir de dados coletados no AVA Moodle.

Dentre os diversos métodos utilizados em EDM, muitos deles originalmente são da área de mineração de dados, como predição, mineração de relações eclusterização, por exemplo. A técnica de agrupamento ou clusterização tem como objetivo principal achar dados que se agrupam naturalmente, classificando os dados em diferentes grupos e/ou categorias. Estes grupos e categorias não são conhecidos a priori. Através de técnicas de agrupamento são automaticamente identificados através da manipulação das características dos dados.

Vários trabalhos envolvendo técnicas de EDM foram desenvolvidos com o intuito de avaliar o comportamento dos usuários de AVAs e propor melhorias na organização de AVAs ou conteúdos dos respectivos cursos. Ricarte e Falci Júnior (2011) aplicaram clusterização utilizando os algoritmos K-means e Mapas Auto-organizáveis (do inglês, Self-Organized Map - SOA) em uma base de *logs* coletada a partir do uso de um AVA na UNICAMP para encontrar grupos de estudantes com comportamento semelhante com o intuito de oferecer retorno a autores e tutores sobre o uso dos conteúdos disponibilizados, bem como oferecer a estudantes sobre seu próprio uso dos recursos do ambiente.

Swedan (2012) aplicou clusterização utilizando K-means para encontrar os diferentes grupos de alunos quanto ao nível de participação no AVA Moodle a partir do uso dos recursos deste. O autor propôs ainda uma ferramenta estatística para os educadores monitorarem os alunos e seus respectivos comportamentos de aprendizagem a partir dos *logs* coletados do Moodle. Analogamente, Bovo et al (2013) utilizaram clusterização para encontrar diferentes grupos de alunos e se estes grupos apresentam diferenças qualitativas e quantitativas.

Romero, Ventura e García (2008) verificaram que as ferramentas de mineração de dados disponíveis são demasiadamente complexas para educadores, que precisam de interfaces mais fáceis de analisar os resultados da mineração para avaliar o processo de aprendizagem. Desta forma é mais provável o administrador do AVA aplicar técnicas de mineração, a fim de produzir relatórios para os educadores que, em seguida, usarão esses relatórios para tomar decisões sobre como melhorar a aprendizagem do aluno e dos cursos *online*.

Neste trabalho a técnica de EDM utilizada é a Clusterização aplicando o algoritmo *K-means* a uma base de *logs* reais coletada do Moodle durante um curso EAD, com o objetivo de encontrar grupos de usuários com comportamento similar para avaliar qualitativamente e quantitativamente estes grupos.

3. Contextualização do problema e descrição da base de dados

No Brasil, dentre as políticas públicas para favorecimento da inclusão digital, têm-se investido na criação e na utilização de centros tecnológicos comunitários, ou telecentros, onde o acesso público às TICs é disponibilizado para as comunidades menos privilegiadas a um custo mínimo ou isento de custos. Neste contexto, além da implantação dos telecentros para acesso à Internet, têm-se a formação de agentes para inclusão digital como aspecto crítico. A proposta envolveu um programa de formação em larga escala que requereu mecanismos de controle e monitoramento para gestores e beneficiadores desse programa, a Rede Nacional de Formação para Inclusão Digital – Rede Telecentros BR [Silva, 2013].

A Rede Telecentros BR até o ano de 2011 contava com a participação de cinco polos regionais (um polo para cada região do País), dois polos estaduais (nos estados de São Paulo e Ceará) e um polo nacional. Sob a responsabilidade dos polos regionais estava a formação dos agentes de inclusão digital (que são os monitores dos telecentros), gestores de telecentros (responsáveis pela administração do telecentro), tutores (que atuavam na formação dos monitores) e supervisores de tutoria (responsáveis pela supervisão e acompanhamento do trabalho dos tutores) [Silva, 2013].

No período de fev/2010 a dez/2012, os membros dos polos de formação se articularam para construir e aplicar o Curso de Formação de Monitores dos Telecentros e a ativação das redes sociais de agentes de inclusão social atuantes nas comunidades. O projeto de formação dos agentes de inclusão digital (monitores) contemplou a oferta de um curso de 480 horas, disponibilizado na plataforma Moodle (Rede Telecentros.BR, 2013) e dividido em dois módulos: 80 horas para uma breve apresentação dos conteúdos da formação e 400 horas com foco específico no desenvolvimento de projetos comunitários e aprofundamento dos conteúdos. Para o desenvolvimento dos projetos comunitários, os agentes de inclusão digital percorreram de acordo com seus interesses e necessidade, sem percurso pré-definido, diferentes temas: comunicação comunitária, redes, cultura digital, comunidade, telecentros. No desenvolvimento do curso, os agentes de inclusão digital contaram com o apoio de tutores e supervisores de tutores dos diversos polos regionais [Silva, 2013].

3.1. Plataforma Moodle

O Moodle é o Ambiente Virtual de Aprendizagem (AVA) selecionado para o desenvolvimento do Curso de Formação de Monitores do Telecentros.BR. A Rede Nacional de Formação para Inclusão Digital produziu o curso de formação de monitores do Telecentros.BR com objetivo de propiciar o desenvolvimento de habilidades no uso de tecnologias da informação e comunicação para dar condições de transformações sociais na comunidade. O curso foi estruturado em dois eixos pedagógicos: acessos a conteúdos e atividades formativas; elaboração e implementação de projetos comunitários. Ofertado em duas fases: a primeira englobava conhecer o ambiente virtual e abordagem prévia do conteúdo que será aprofundado na fase dois que aborda principalmente o projeto comunitário, onde cada tema estudado poderá servir de apoio para realização de ações com a comunidade. Entre os temas abordados: telecentros, comunidade, comunicação comunitária, inclusão digital, redes e cultura digital.

No ambiente virtual de aprendizagem – Moodle foi organizado o curso para formação dos monitores em: Fase 1: Ambientação e Voo Rasante; Fase 2: Projeto Comunitário. Também foi necessário realizar um curso para formação de tutores que atuaram auxiliando e orientando os monitores durante a formação do telecentros.br. No

curso dos tutores foi abordado educação a distância, articulação social e inclusão digital. Os tutores foram selecionados com base no conhecimento sobre tecnologia, valorizando a experiência em inclusão digital.

O processo de formação foi realizado em grupo de “n” monitores para um tutor, que fizeram o acompanhamento para o primeiro acesso a plataforma, incentivaram a participação no ambiente virtual e no desenvolvimento do projeto comunitário, assim como realizaram a avaliação dos monitores mensalmente. Os tutores tiveram o acompanhamento e auxílio dos supervisores de tutoria no processo de formação. Os supervisores eram pessoas que fizeram parte dos polos regionais, facilitadores da formação, acompanhavam os tutores e monitores no processo de formação, realizaram também avaliação das turmas com base nas informações levantadas na plataforma moodle e com as informações repassadas pelos tutores. A coordenação pedagógica realizou acompanhamento dos supervisores e tutores com base em informações obtidas na plataforma e no sistema de avaliação.

4. Experimentos realizados

Para a realização deste trabalho foi utilizada a base de dados do Moodle utilizado na Rede Telecentros BR por aproximadamente mil e quatrocentos alunos (monitores) no curso de formação para inclusão digital promovido pela Rede Nacional de Formação para Inclusão Digital, a qual foi responsável pela construção, manutenção e aplicação do referido curso.

A coleta de dados ocorreu de maneira implícita através do uso e interação dos monitores na plataforma Moodle e armazenamento na base de dados da plataforma. O Moodle armazena todas as interações dos usuários na plataforma em forma de logs: cada clique que o usuário realiza na plataforma para fins de navegação, bem como detalhes das atividades que os estudantes participaram. Uma vantagem de registrar as atividades de um usuário em uma plataforma em forma de log é que grande grandes volumes de dados podem ser armazenados automaticamente [Rogers, Sharp and Preece, 2011].

O Moodle armazena os logs em uma base de dados relacional, com cerca de 145 tabelas inter-relacionadas, porém neste trabalho as informações utilizadas encontravam-se somente na tabela *mdl_log* responsável pelo armazenamento de eventos do sistema. O pré-processamento consistiu na limpeza das informações das demais tabelas do Moodle em conjunto com o filtro para eliminar alunos excluídos e administradores do sistema procurando mitigar a interferência destes grupos de usuários.

Neste trabalho foi utilizado o algoritmo *K-means* que é um método de clusterização (organização dos objetos similares, em algum aspecto, em um grupo) pertencente à classe dos algoritmos de aprendizados de máquina. Tem como finalidade dividir um determinado número de objetos em áreas chamadas de *cluster*. Estas áreas são constituídas por uma coleção de objetos que são similares entre si e diferentes dos objetos pertencentes aos outros *clusters*. A distância de cada objeto para cada *cluster* vai determinar em que *cluster* o objeto ficará alocado. Cada objeto pertence ao *cluster* no qual possui o elemento central (centróide) mais próximo deste objeto

Apartir da etapa de pré-processamento a clusterização é permitida usando o algoritmo K-means para duas dimensões onde são tratadas as colunas: “course” e “userid”, contando o número de acessos de cada usuário para cada curso, também são filtrados somente eventos dentro da categoria “course”. Após a composição do *dataset* para o algoritmo de clusterização o mesmo será executado para: $k=5$, pois são estabelecidos pelo programa quatro níveis de avaliações qualitativas: Excelente, Bom,

Regular e Insuficiente, contudo foi acrescentada uma classe para que *outliers*(valores fora do esperado) possam ser minimizados nas outras classes (FAYYAD, U et al, 1996). Os resultados do *K-means* são grupos definidos que passam pela análise de quantitativos de alunos, acessos e cursos.

A partir do mesmo é possível inferir características comuns entre grande número de usuários no mesmo grupo e ate mesmo o grupo dos usuários mais experientes na plataforma, chamados monitores.

5. Resultados

Para o pré-processamento o número de linhas da base foram resumidas para 13283 linhas, pois as linhas de log ficaram restritas as características citadas na metodologia. É importante ressaltar o tempo de execução em 5 minutos e 55 segundos, para o pré-processamento e K-means. Todos os procedimentos foram executados no mesmo computador com os seguintes requisitos de hardware: Processador Core i7-3632QM @ 2.20Ghz, Memória de 8GB @1333Mhz e sistema operacional Linux distribuição Ubuntu 14.04.

Para o K-means foram encontrados os 5 clusters (denominados K1, K2, K3, K4, K5) com as seguintes características como mostra aTabela1:

Tabela 1. Clusters encontrados

	K1	K2	K3	K4	K5
Número de Usuários	5	32	183	825	3251
Acessos	228728	598867	1197017	2024758	1152300
Cursos	50	50	50	49	50
Acesso Médio por Usuário	45745,6	18714,59375	6541,076502732	2454,252121212	354,44478622
Acesso Médio por Usuário e Por Curso	914,912	374,291875	130,821530055	50,086777984	7,088895724
Ação Mais Utilizada	<i>View</i> 93930 execuções	<i>View</i> 266186 execuções	<i>View</i> 614599 execuções	<i>View</i> 1183165 execuções	<i>View</i> 93930 execuções
Recurso Mais Utilizado	<i>ErrorLogin</i> 59430 acessos	<i>Course</i> 153197 acessos	<i>Forum</i> 294337 acessos	<i>Course</i> 513217 acessos	<i>Course</i> 310376 acessos
Quantidade Caminhos Médios	145	163	167	137	133

Os resultados apresentados na Tabela1 demonstram que a partir da clusterização do log da plataforma Moodle foram encontrados grupos distintos de usuários. É possível reconhecer com base nestes resultados a distância entre o maior e menor cluster, isso ocorre, pois no caso do cluster K1, pelo pequeno número de usuários e grande número de acessos com erro no Login, o grupo demonstra ser formado provavelmente por usuários com comportamento fora do esperado (*outliers*). No cluster K2, identificamos pelo número de usuários e a grande quantidade de acessos, os usuários responsáveis pela capacitação dos monitores, chamados Tutores. Nos clusters K3, K4 e K5 encontramos números de usuários distintos, porém com a média de acesso geral

semelhante e com a média por usuário decrescente, respectivamente, então, é possível inferir que estão localizados alunos, chamados monitores, com bom acesso ao sistema no cluster K3, monitores com acesso regular ao sistema no cluster K4 e monitores com um baixo número de acesso na plataforma no cluster K5.

5.1. Discussão dos Resultados

A partir dos clusters encontrados, buscou-se fazer uma análise estatística dos recursos mais utilizados em comum pelos usuários de cada grupo. Entende-se por caminho médio quais as ações que foram mais utilizadas junto a recursos na plataforma por cada *cluster* encontrado, estes caminhos médios foram levantados em busca de encontrar as ações em comum dos usuários de cada cluster na plataforma.

Partindo dos resultados mostrados na Tabela 1, podemos inferir que os monitores do cluster K3 realmente foram mais participativos, pois o recurso mais utilizado por estes foi o *Forum*, como mostra o gráfico dos recursos mais utilizados pelo cluster K3 na Figura 1. Os monitores do cluster K4 tiveram uma maior participação nos materiais dos seus respectivos cursos, como mostra o gráfico estatístico do uso de recursos apresentado na Figura 2. Os usuários do cluster K5 também tiveram participação razoável em seus respectivos cursos, como apresenta a Figura 3, com ampla utilização dos materiais dos seus respectivos cursos.

Percebe-se que o recurso que mais utilizado na plataforma pelos alunos com acesso regular e com baixo acesso, clusters K4 e K5, durante a formação foi o “Curso”, seguido pelo “Recursos” (Arquivos, Mídias, etc.). Enquanto que o recurso de “Fórum” foi o mais utilizado pelos alunos com maior número de acessos à plataforma, seguido pelo “Curso”. Ferramentas importantes de comunicação como Chat, por exemplo, foram muito pouco utilizadas por estes alunos.

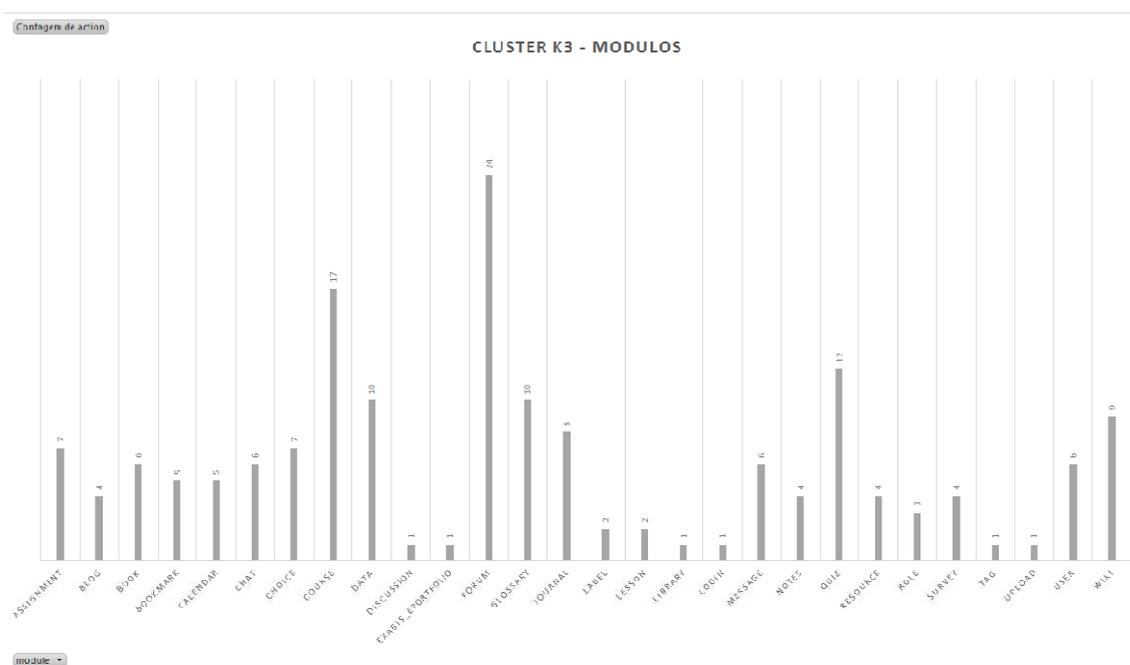


Figura 1. Uso dos recursos pelo cluster K3

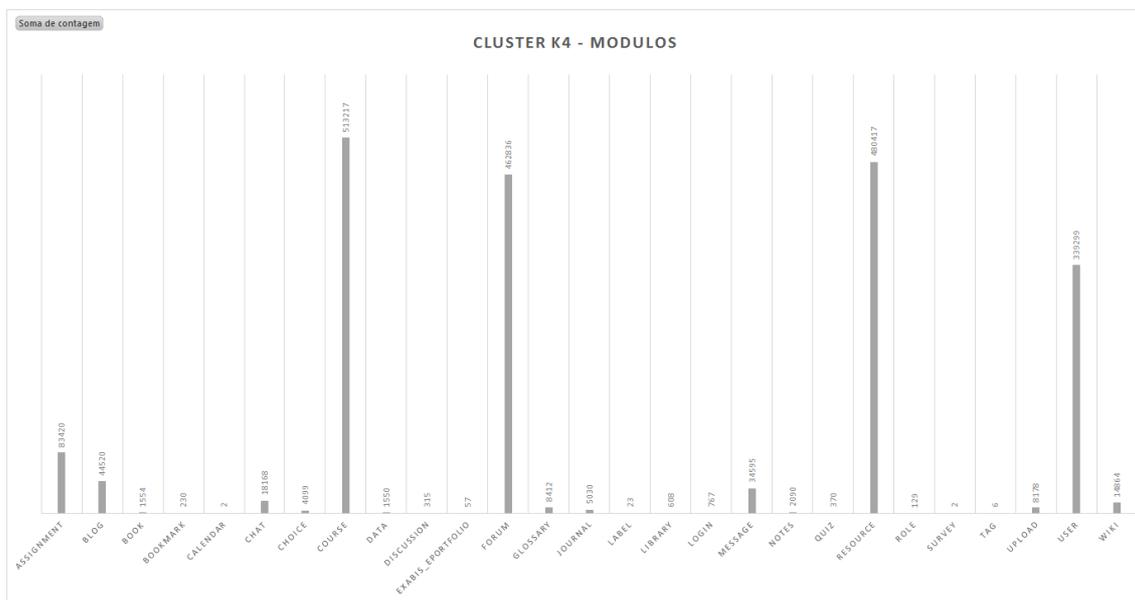


Figura 2. Uso dos recursos pelo cluster K4

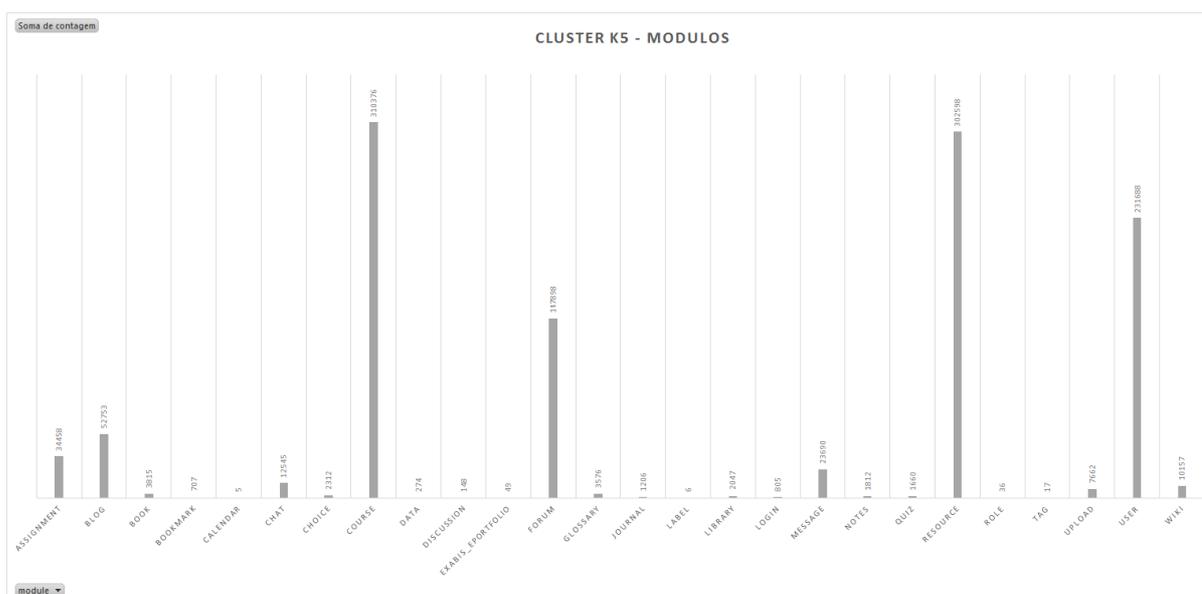


Figura 3. Uso dos recursos pelo cluster K5

5. Conclusões e trabalhos futuros

Este trabalho apresentou como proposta exploratória (Gil, 2008) a utilização de técnicas de EDM com o intuito de encontrar grupos de alunos de um programa educacional de larga escala através da análise de log de uma plataforma de Educação a Distância. Com base nas técnicas que são reportadas pela literatura, aplicou-se o algoritmo *K-means* em busca do agrupamento de alunos com comportamentos semelhantes na plataforma utilizando a análise do caminho médio para encontrar as ações em comum dos usuários de cada *cluster* dentro da plataforma Moodle.

Os resultados encontrados são de extrema importância para o gerenciamento dos cursos no Moodle e para as análises quantitativas e qualitativas feitas no programa Telecentros BR, pois o governo poderá verificar e avaliar quais os recursos que mais têm contribuído na formação dos monitores, bem como melhorar o processo de capacitação dos mesmos através da análise do perfil dos alunos da programa e até mesmo através de metodologias de ensino personalizadas para cada grupo encontrado. É possível identificar desde as primeiras atividades na plataforma o perfil do aluno, assim o tutor pode adaptar e/ou modelar o trabalho pedagógico realizado objetivando alcançar alunos um melhor desempenho dos alunos durante a capacitação, identificando os alunos com um baixo rendimento de aprendizagem buscando modificar incentivar e melhorar o processo de aprendizagem destes.

Como trabalho futuro pretende-se correlacionar os grupos de monitores (alunos) encontrados neste trabalho com seus respectivos caminhos médios e com suas respectivas avaliações realizadas em outro sistema pelos tutores (educadores). Para tal será necessário a aplicação técnicas de KDD utilizando algoritmos de Mineração de Texto, para a obtenção das notas dos monitores, pois as avaliações são qualitativas e não quantitativas. Conseqüente pretende-se utilizar outras técnicas de Inteligência Artificial para classificação dos alunos segundo suas notas.

Referências

- Baker, R.S.J.D. (2010) Data Mining for Education. In: McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education. Oxford, UK: Elsevier, 3ed.
- Baker, R.S.J.D., Isotani, S. and, de Carvalho, A.M.J.B. (2011) “Mineração de Dados Educacionais: Oportunidades para o Brasil”. Revista Brasileira de Informáticsa Educação, vol. 19, no. 2, p. 2-13.
- Bovo, A., Sanchez, S., Héguy, O. and Duthen, Y. (2013) “Clstering Moodle Data as a Tool for Profiling Students”, In: Second International Conference on e-Learning and e-Technologies in Education (ICEEE), p. 121-126
- Capterra (2014). The Top 20 Most Popular Lms Soft Ware Solutions. [Http://Www.Capterra.Com/Top-20-Lms-Software-Solutions.Uzjrp7wnsp](http://www.capterra.com/top-20-lms-software-solutions.uzjrp7wnsp).
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996) From Data Mining to Knowledge Discovery: An Overview, in Advances in Knowledge Discovery and Data Mining, R. Uthurusamy, eds., MIT Press, Cambridge, Mass., pp. 1-36.
- Frascareli, A.M.F. and Pimentel, E.P (2012) “Aplicando Técnicas de Bibliometria, Mineração de Texto e Vizualização na Identificação de Temas e Tendências de Pesquisa em *e-Learning*”, Anais do XXIII SBIE, p. 26-30.
- Gil, A. C. (2008) “Métodos e Técnicas de Pesquisa Social”, São Paulo, Atlas.
- Gosain, A and Kumar, A, (2009) “Analysis of health care data using different data mining techniques”, International Conference on Intelligent Agent & Multi-Agent Systems (IAMA), vol. 1, no. 6, pp. 22-24
- Gottardo, E. Kaestner, C. and Noronha (2012) “Previsão de Desempenho de Estudantes em Cursos EAD Utilizando Mineração de Dados: Uma Estratégia Baseada em Séries Temporais”, Anais do XXIII SBIE.

- Hu, X. (2011) “Data mining and its applications in bioinformatics: Techniques and methods”, IEEE International Conference on Granular Computing (GrC), vol.3 (3): , no. 3, p. , 8-10.
- Klösgen, W. and Zytow, J. (2002) Handbook of data mining and knowledge discovery. New York: Oxford University Press.
- Kovalerchuk, B. and Vityaev, E. (2005) “Data Mining for Financial Applications”, Data Mining and Knowledge Discovery Handbook. Springer US, p. 1203-1224.
- Mostow, J. and Beck, J. (2006) Some useful tactics to modify, map and mine data from intelligent tutors. Journal Natural Language Engineering, vol. 12, pp. 195-208.
- Mostow, J., Et Al. (2005) An educational data mining tool to browse tutor–student interactions: Time will tell! In: Proceedings of the workshop on educational data mining, pp. 15–22.
- Okonkwo, R.O. and Enem, F.O. (2011) “Combating Crime and Terrorism Using Data Mining Techniques”, 10th International Conference Information Technology for People-Centered Development (ITePED).
- Ricarte, I. L. M., Falci Junior, G. R. (2011) “A Methodology for Mining Data from Computer-Supported Learning Environments”, Informática na Educação: teoria & prática, Porto Alegre, v. 14, n. 2, p. 83-94.
- Rogers, Y., Sharp, H., and Preece, J. (2011) Interaction Design: Beyond Human - Computer Interaction. Wiley, 3rd ed.
- Romero, C., Ventura, S. and García, E. (2008) “Data mining in course management systems: Moodle case study and tutorial”, Computers & Education 51 (1), 368-384, 405.
- Silva, A., et al. (2013) Análise de Redes Sociais para avaliação e monitoramento de programas de treinamento em larga escala baseados no uso de ambientes de aprendizagem e redes sociais online. II Brazilian Workshop on Social Network Analysis and Mining.
- Swedan, M.I. (2012) “Students Learning Behavior in Moodle System Using DataMining Techniques”, In: Universal Journal of Applied Computer Science and Technology, 2 (4): 318-823.
- Wicker, P. and Breuer, C. (2010) “Analysis of problems using Data Mining techniques – findings from sports clubs in Germany”, In: European Journal for Sport and Society, vol. 7 (2), p.: 131-140
- Witten, I.H., Frank, E. and Hall, M. A. (2011) Data Mining: Practical Machine Learning Tools and Techniques, San Francisco: Morgan Kaufmann, 3ed.