

Mineração de dados educacionais e Mundos Virtuais: um estudo exploratório no OpenSim

Felipe Becker Nunes¹, Gleizer Bierhalz Voss^{1,2}, Silvio César Cazella^{2,3}

¹Universidade Federal do Rio Grande do Sul (UFRGS)

Av. Paulo Gama, 110, Porto Alegre, RS, Brasil - CEP 90040-060

²Instituto Federal Farroupilha (IF Farroupilha) – Campus São Vicente do Sul
Rua Vinte de Setembro, S/N, São Vicente do Sul, RS, Brasil - CEP 97420-000

³Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)

Rua Sarmiento Leite, 245 - Porto Alegre, RS, Brasil - CEP 90050-170

{nunesfb, silvio.cazella}@gmail.com, gleizer.voss@iffarroupilha.edu.br

Abstract. *Use of virtual world is expanding in the educational scene, as well as the range of research opportunities related to these. This paper aims to explore the feasibility of implementing educational data mining (EDM) in virtual worlds, to identify possible patterns of users and allow changes in educational planning. A case of study was performed with a virtual laboratory of chemistry in OpenSim, where interactions were simulated with synthetic data and analyzed using association rules with Apriori algorithm. The results demonstrated the feasibility of the proposal, being possible to use EDM within virtual worlds to identify user behavior patterns.*

Resumo. *O uso dos mundos virtuais vem se expandindo no cenário educacional, assim como a gama de possibilidades de pesquisas relacionadas a estes. Este artigo buscou explorar a viabilidade da aplicação de mineração de dados educacionais (MDE) nos mundos virtuais, para identificar possíveis padrões dos usuários e permitir alterações no planejamento pedagógico. Um estudo de caso foi realizado com um laboratório virtual de química no OpenSim, em que foram simuladas interações com dados sintéticos e analisados por meio da tarefa de regras de associação com o uso do algoritmo Apriori. Os resultados demonstraram a viabilidade da proposta, sendo possível o uso de MDE dentro dos mundos virtuais para identificar padrões de comportamento dos usuários.*

1. Introdução

A evolução dos recursos tecnológicos no decorrer das últimas décadas tem acarretado em mudanças nos paradigmas tradicionais de ensino e suas formas de aplicação. A introdução das Tecnologias de Informação e Comunicação (TIC) tem proporcionado novas possibilidades de uso dos recursos computacionais como elemento de apoio e mesmo motivação no processo de ensino e aprendizagem, tanto em ambientes presenciais, quanto no ensino à distância.

Dentre as alternativas que têm figurado no cenário educacional, está o paradigma de Ambientes Virtuais Imersivos ou Mundos Virtuais (MV). Garrido et al. (2010) explicam que os Mundos Virtuais representam mundos sociais persistentes, gerados por meio de ambientes computacionais simulados que permitem a interação entre avatares. Este paradigma se apresenta como um recurso interessante para a construção de conteúdos transdisciplinares, não lineares e dinâmicos sobre um mundo virtual, que pode ser disponibilizado no próprio computador do aluno.

Os movimentos e as iterações dos avatares podem ser monitorados e informações podem ser analisadas para obter padrões de comportamento. Devido a grande quantidade de dados resultantes das interações dos usuários nos MV, abre-se uma nova gama de possibilidades por intermédio do uso de técnicas de Mineração de Dados Educacionais (MDE).

Segundo Romero e Ventura (2010), a área de MDE está preocupada com o desenvolvimento, pesquisa e aplicação de métodos computadorizados para detectar padrões em grandes conjuntos de dados educacionais, que de outra forma seriam difíceis ou impossíveis de analisar devido ao enorme volume de dados. No escopo dos MV, no que concerne ao tratamento e processamento das pilhas de dados coletadas a partir das interações dos usuários, o uso da MDE visa possibilitar a interpretação de informações que possam auxiliar no planejamento pedagógico das atividades no mundo virtual.

Com base nas asserções proferidas, o presente estudo tem por objetivo explorar a aplicabilidade de técnicas de MDE em MV. Para tanto, um estudo de caso foi desenvolvido em um laboratório de química criado no OpenSimulator (OpenSim), em que foram coletados dados sintéticos (simulação de usuários) referentes às preferências de materiais e nível de conhecimento (*expertise*). Assim, foram analisados possíveis padrões de comportamento na base de dados por meio da tarefa de mineração de dados denominada Regra de Associação com o uso do algoritmo Apriori, que possibilitou a identificação de eventuais mudanças no planejamento pedagógico das atividades.

O restante deste artigo está organizado da seguinte forma: na seção 2 são apresentados os trabalhos correlatos; nas seções 3 e 4 são abordados os conceitos de Mundos Virtuais e Mineração de Dados; na seção 5 é descrita a metodologia utilizada nesta pesquisa; na seção 6 é realizada a descrição do experimento realizado; e por fim, na seção 7 estão as considerações finais deste trabalho.

2. Trabalhos Relacionados

Apesar de relevante e atual, poucas propostas para a utilização de MDE no contexto dos MV têm sido desenvolvidas com o objetivo de melhorar o processo de ensino e aprendizagem. Alguns desses trabalhos foram analisados a fim de utilizar o conhecimento adquirido a partir da experiência dos mesmos para a realização deste estudo exploratório (Quadro 1).

O trabalho de Orgaz et al. (2012) é o que apresenta maior similaridade com o estudo apresentado neste artigo. Os autores desenvolveram uma clusterização dos comportamentos dos avatares com base nas suas interações no mundo virtual. No entanto, diferente da proposta deste trabalho, além da técnica utilizada (algoritmo K-

means), a mineração foi realizada com base nas interações via *chat* que ocorreram no mundo virtual.

Já os trabalhos de Fernández-Gallego et al. (2013) e Donalek et al. (2014) têm objetivos distintos deste artigo. Enquanto o primeiro objetiva descobrir fluxos de aprendizagem com base na proposta de um *framework* para *Learning Analytics* (LA), o segundo é voltado para a visualização de dados efetivos dentro de um MV, para isso são utilizadas tecnologias emergentes (e.g., Oculus Rift™) associadas ao conceito de Realidade Virtual Imersiva.

Quadro 1. Comparativo entre trabalhos relacionados

Autores / Ano	Objetivo do trabalho	Técnica utilizada
Orgaz et al. (2012)	Clusterização de avatares	Algoritmo K-means
Fernández-Gallego et al. (2013)	Visualização de dados efetivos	Realidade Virtual Imersiva (apenas visualização)
Donalek et al. (2014)	Proposta de um framework para learning analytics	Formalismo - Redes de Petri

Diferentemente dos trabalhos apresentados, o presente artigo utiliza a tarefa de regras de associação por meio da aplicação do algoritmo Apriori, que realiza buscas sucessivas em toda a base de dados, mantendo um ótimo desempenho em termos de tempo de processamento (AGRAWAL e SRIKANT, 1994). Assim, possibilitando aos estudantes uma experiência individualizada e diferenciada de aprendizagem.

3. Mundos Virtuais

Para Bainbridge (2010), os Mundos Virtuais podem ser definidos como ambientes *online* persistentes, gerados por computador onde as pessoas podem interagir, seja para o trabalho ou lazer, de forma comparável ao mundo real. Como principais exemplos de Mundos Virtuais que têm sido utilizados por diferentes pesquisadores estão: OpenSim, Second Life e OpenWonderland.

A natureza imersiva deste tipo de ambiente incentiva os alunos (como avatares) a executarem atividades de aprendizagem que não estavam inicialmente previstas pelos professores, como interagir com outros alunos através de *chats*, procurar novos conteúdos multimídia ou utilizar novos artefatos físicos 3D para aprender sobre um determinado tópico (FERNÁNDEZ-GALLEGO et al., 2013). Esta autonomia fornecida aos estudantes pode auxiliar no desenvolvimento de suas ações próprias e investigações acerca de novos conhecimentos calcados nos diversos recursos presentes em um ambiente 3D.

Diferentes tipos de informações podem ser coletadas no MV, a partir dos avatares, as quais envolvem a sua posição atual, campo de visão, e, outros tipos de interações, como conversas ou operações realizadas no ambiente (objetos de construção, resultados de *scripts* de compilação, etc.) (ORGAZ et al., 2012). Dada a grande quantidade de dados gerados nas interações dos usuários e a ausência de uma base de dados mais organizada e preparada para a análise de informações, esta tarefa se caracteriza como complexa e desafiadora.

O uso de técnicas de mineração de dados se torna possível mediante a preparação da base (pré-processamento), de forma que possam ser extraídos dados

relevantes e padronizados para inserção nas ferramentas de mineração e posterior interpretação das informações obtidas. Desta forma, a próxima seção apresenta uma breve explanação da área de MDE e sua aplicabilidade no escopo deste trabalho.

4. Mineração de Dados Educacionais

Kampff et al. (2014) explicam que a partir do desenvolvimento da MDE é possível desenvolver mecanismos e ferramentas educacionais mais eficientes, modelos para identificar alunos com dificuldades, aperfeiçoar os materiais didáticos e desenvolver métodos pedagógicos mais eficazes. A partir da análise dos dados referentes às interações dos alunos, diferentes informações podem ser obtidas, como perfis e padrões relevantes ao planejamento futuro das atividades educacionais.

Os dados coletados para análise não estão restritos às interações individuais dos alunos com um sistema educacional (comportamento de navegação, respostas em questionários e exercícios interativos), mas pode também incluir dados de atividades colaborativas (*chat* de texto), dados administrativos (escola, distrito, professor), dados demográficos (sexo, idade, notas escolares), a afetividade do aluno (e.g., motivação, estados emocionais), entre outros (ROMERO e VENTURA, 2010). Portanto, as etapas envolvidas para análise e interpretação dos dados por meio de técnicas de mineração se configuram como essenciais para que todo processo ocorra de forma correta e gere resultados significativos.

Em um ambiente virtual, cuja interação se dá de forma não presencial, como é o caso dos MV, diversos aspectos relacionados ao comportamento dos alunos permeiam uma análise mais exigente por parte do professor. Isso acaba por gerar uma grande quantidade de dados a ser tratada e analisada, o que exige a utilização de ferramentas e técnicas para colaborar no processo descritivo e interpretação das informações.

Desta forma, no escopo deste trabalho foi selecionado o uso da tarefa de regras de associação, que segundo Grawal et al. (1993), representa um padrão de relacionamento entre itens de dados do domínio da aplicação que ocorre com uma determinada frequência na base de dados. O algoritmo de associação denominado Apriori foi aplicado. Conforme Agrawal e Srikant (1994), esse algoritmo serve para identificar conjuntos de itens frequentes e regras de associação de aprendizagem, buscando identificar os itens individuais mais frequentes e estendê-los para conjuntos de itens maiores, desde que esses conjuntos de itens apareçam suficientemente vezes no banco de dados. Aqueles conjuntos de dados que forem mais frequentes podem ser usados para determinar as regras de associação e tendências da base analisada.

Com base nisso, este estudo buscou identificar possíveis padrões de comportamentos relacionados às preferências de materiais e *expertise* dos usuários, tendo como base dados sintéticos gerados a partir da simulação de interações dos usuários no MV. A próxima seção detalha os procedimentos adotados para o desenvolvimento deste artigo.

5. Metodologia

O delineamento da pesquisa desenvolvida neste trabalho pode ser classificada como estudo de caso (YIN, 2005). Para alcançar os objetivos propostos na seção 1, após um

levantamento teórico acerca dos principais tópicos abordados neste trabalho, foram seguidas as seguintes etapas:

A primeira ficou caracterizada pela definição da infraestrutura tecnológica utilizada para aporte desta pesquisa. O OpenSim 0.8.1 foi escolhido por ser gratuito, *open source* e com uma vasta documentação no meio acadêmico. Para a visualização do MV foi selecionado o *viewer* Singularity, por se adequar às necessidades identificadas na construção, visualização e criação de objetos no MV. Para a hospedagem e funcionamento das ferramentas utilizadas, o Wamp Server foi escolhido por ser gratuito e conter três recursos acoplados à sua instalação (PHP, MySQL e Apache).

A segunda etapa abordou o estudo de caso em si, no qual um laboratório virtual de química desenvolvido no OpenSim foi utilizado (Nunes et al., 2014), o qual contém diferentes tipos de atividades educacionais, como a apresentação de *slides*, vídeos e atividades de cunho prático demonstrando como a densidade de diferentes elementos ocorrem através de simulações. Neste laboratório, variados tipos de dados são gravados resultantes das interações dos usuários, sendo armazenados na tabela “Activities_Records”, que continha 10 atributos (detalhados na Seção 5).

Por fim, a terceira etapa compreendeu a análise dos dados resultantes das interações ocorridas no MV. Foram gerados 464 registros a partir da ferramenta *online* GenerateData¹, a partir de uma opção aleatória por parte dos autores, sem haver uma especificidade para tal número. A escolha por popular a base com dados sintéticos ocorreu devido à elevada quantidade de registros necessários para inferir técnicas de mineração de dados e também para viabilizar a proposta apresentada neste artigo, visto que para a interação com usuários reais, um extenso período de tempo se faria necessário.

Quanto à tarefa de mineração de dados, a Regra de Associação com o uso do algoritmo Apriori foi aplicada. Em particular, este algoritmo possui o fluxo de operação requerido para o escopo deste trabalho, com a identificação de padrões de comportamento e possíveis perfis de usuário, com a mensuração dos elementos que mais aparecem na base de dados repetidamente. A ferramenta Weka², também de licença gratuita e vasta documentação no meio acadêmico, foi escolhida para a análise dos dados gerados e interpretação de possíveis padrões identificados. A próxima seção apresenta o estudo de caso desenvolvido neste trabalho, no qual são detalhados os recursos existentes no laboratório virtual de química e a interpretação dos resultados obtidos com as técnicas de MDE aplicadas.

6. Estudo de caso

O estudo de caso desenvolvido nesta pesquisa visou analisar os dados gerados a partir de simulações das interações de usuários dentro de um laboratório virtual de química no OpenSim. O laboratório de química conta com cinco diferentes divisões: sala de entrada, sala de conteúdos, sala de vídeos, sala de experimentos e sala de atividades

¹ Site oficial. Disponível em: <http://generatedata.com/>

² Site oficial. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>

avaliativas. Na sala de entrada, o usuário realiza sua identificação (idade e gênero), e, responde ao pré-teste disponibilizado para identificação de sua *expertise* inicial.

A sala de conteúdos contém *slides*, textos e imagens acerca dos conteúdos trabalhados no laboratório e que estão hospedados em *websites*, sendo apresentados aos avatares quando esses clicam no objeto para visualizar. Da mesma forma, na sala de vídeos, diversos filmes podem ser vistos pelos usuários como parte do conteúdo base para entendimento dos experimentos existentes no laboratório (NUNES et al., 2014).

Na sala de experimentos (Figura 1), três tipos de simulações foram construídos para ser visualizadas pelos usuários, sendo: densidade de diferentes tipos de líquidos, entre água e cobre, e acerca de águas com temperaturas diferentes. As ações nestes objetos são executadas com o uso de scripts inseridos nos objetos, que ao serem clicados, realizam as movimentações previamente estabelecidas, a troca de texturas ou aparência dos objetos e emite a descrição detalhada de cada etapa que ocorre durante o experimento (NUNES et al., 2014).



Figura 1. Sala de experimentos do laboratório

Efetuada a descrição dos recursos presentes no laboratório virtual, se faz necessário detalhar a forma com que foram organizados os dados para inserção no Weka e posterior aplicação das técnicas de mineração de dados. Seguindo as orientações propostas por Fayyad et al. (1996), o processo de Descoberta de Conhecimento em Bases de Dados (DCBD) é composto por cinco etapas, sendo as seguintes: Seleção, Pré-Processamento, Transformação, Mineração de Dados e Interpretação/Avaliação. O resultado final obtido é o conhecimento acerca da problemática que foi proposta inicialmente a ser solucionada e respondida.

Inicialmente foi definido o escopo do trabalho, o qual pretendia analisar possíveis padrões existentes com relação às preferências por materiais (vídeos, slides, textos, imagens) e a *expertise* dos usuários (inicial e final) dentro do MV. Assim, poderiam ser inferidas classificações de perfis dos usuários em torno destes elementos, que ajudariam a reunir informações importantes para o processo de elaboração das atividades pedagógicas no MV.

Dentre os diversos dados gerados das interações dos usuários nos MV, o presente estudo buscou reunir um conjunto de dados específicos para o processamento e aplicação das técnicas de mineração. Desta forma, uma tabela específica foi criada com 10 atributos que contém elementos armazenados em outras tabelas do banco de dados do OpenSim, assim como novos atributos.

A descrição dos atributos contidos na tabela “Activities_Records” é detalhada a seguir:

- ID: identificação única de cada registro na tabela;
- Nome do usuário (first_name): contém o nome dado ao avatar no mundo virtual;
- Idade (age) e gênero (gender): no pré-teste são coletados estes dados para complementar a análise do perfil dos usuários;
- Data do Último Acesso (date_time): data em que ocorreu o último acesso ao Mundo Virtual;
- Preferências de Materiais (type_material): pode ser Vídeo, *Slide*, Texto ou Imagem. São coletadas a partir de questionário e acesso ao material dentro do MV;
- Acesso à sala de experimentos (access_experiment_room): uma *flag* (T ou F) para saber se houve acesso a esta sala e interação com os experimentos;
- Nível inicial e final de *expertise* (level_expertise): pode ser Básico, Intermediário ou Avançado. É coletado por meio de pré-teste (nível inicial), e testes durante a interação do usuário com o MV (nível final).

Vale ressaltar que a inserção de dados nesta tabela a partir das interações ocorridas no MV foi implementada por meio do uso de *scripts* OpenSim Script Language (OSSL), sendo realizados testes manuais com avatares no OpenSim para confirmar o correto funcionamento da proposta.

O funcionamento ocorre por meio de dois arquivos, onde o primeiro é o *script* inserido no MV para capturar interações previamente definidas e enviar para uma página PHP (segundo arquivo), que recebe as informações e retorna a confirmação de gravação do registro do usuário. Concomitantemente ao retorno da mensagem, a página PHP realiza a inserção dos registros na tabela “Activities_Records”.

Com isso, foram realizados os processos de limpeza dos dados e redução no banco de dados do OpenSim, reduzindo a uma tabela específica com o objetivo de obter uma melhor extração do conhecimento. Desta forma, para efeitos de testes, foi necessária a população desta tabela com dados sintéticos gerados a partir da ferramenta *online* GenerateData. Um total de 464 registros foram adicionados na tabela criada no banco de dados OpenSim, sendo efetuada a sua exportação para organização de acordo com o padrão Attribute-Relation File Format (ARFF), exigido para importação na ferramenta Weka.

A próxima etapa consistiu na seleção do algoritmo de mineração a ser aplicado nos dados coletados. Com a importação efetuada no Weka, foi aplicada a técnica de associação por meio do algoritmo Apriori, para obter informações acerca de possíveis padrões existentes relacionados ao perfil dos usuários.

Foram definidos os seguintes atributos para ser analisados: “gender”; “type_material_one”, “type_material_two”, “inicial_level_expertise” e “final_level_expertise”. Os atributos “id”, “first_name”, “age”, “date_time” e “access_experiment_room” foram ignorados por não serem considerados relevantes para esta análise em específico. Porém, podem ser utilizados para análises mais complexas no decorrer do processo. A configuração do algoritmo Apriori foi definida

como a seguinte: “Apriori -N20 -T0 -C0.9 -D0.05 -U1.0 -M0.1 -S-1.0 -c-1”. A Figura 2 apresenta as regras de associação obtidas com a aplicação do algoritmo.

```

Associator output
Generated sets of large itemsets:

Size of set of large itemsets L(1): 16

Size of set of large itemsets L(2): 23

Size of set of large itemsets L(3): 8

Best rules found:

1. final_level_expertise=ADVANCED 266 ==> inicial_level_expertise=BASIC 266   conf:(1)
2. inicial_level_expertise=BASIC 266 ==> final_level_expertise=ADVANCED 266   conf:(1)
3. final_level_expertise=BASIC 257 ==> inicial_level_expertise=ADVANCED 257   conf:(1)
4. inicial_level_expertise=ADVANCED 257 ==> final_level_expertise=BASIC 257   conf:(1)
5. final_level_expertise=MEDIUM 254 ==> inicial_level_expertise=MEDIUM 254   conf:(1)
6. inicial_level_expertise=MEDIUM 254 ==> final_level_expertise=MEDIUM 254   conf:(1)
7. type_material_two=SLIDE 195 ==> type_material_one=TEXT 195   conf:(1)
8. type_material_one=TEXT 195 ==> type_material_two=SLIDE 195   conf:(1)
9. type_material_two=VIDEO 195 ==> type_material_one=IMAGE 195   conf:(1)
10. type_material_one=IMAGE 195 ==> type_material_two=VIDEO 195   conf:(1)
11. type_material_two=TEXT 194 ==> type_material_one=SLIDE 194   conf:(1)
12. type_material_one=SLIDE 194 ==> type_material_two=TEXT 194   conf:(1)
13. type_material_two=IMAGE 193 ==> type_material_one=VIDEO 193   conf:(1)
14. type_material_one=VIDEO 193 ==> type_material_two=IMAGE 193   conf:(1)
15. gender=FEMALE final_level_expertise=MEDIUM 134 ==> inicial_level_expertise=MEDIUM 134   conf:(1)
16. gender=FEMALE inicial_level_expertise=MEDIUM 134 ==> final_level_expertise=MEDIUM 134   conf:(1)
17. gender=MALE final_level_expertise=ADVANCED 133 ==> inicial_level_expertise=BASIC 133   conf:(1)
18. gender=MALE inicial_level_expertise=BASIC 133 ==> final_level_expertise=ADVANCED 133   conf:(1)
19. gender=FEMALE final_level_expertise=ADVANCED 133 ==> inicial_level_expertise=BASIC 133   conf:(1)
20. gender=FEMALE inicial_level_expertise=BASIC 133 ==> final_level_expertise=ADVANCED 133   conf:(1)

```

Figura 2. Aplicação do algoritmo Apriori no Weka

A análise dos dados apontou para 20 regras que apresentavam os padrões encontrados, nos quais houve o cruzamento relacional entre o tipo de material, gerando associações com a preferência por tipo de material primário e secundário, entre *expertise* inicial e final, bem como a relação com o gênero. A interpretação dos resultados obtidos demonstra que, em uma hipotética simulação das interações dos usuários com as atividades educacionais dispostas no laboratório virtual de química no OpenSim, poderiam ser inferidas as seguintes asserções:

- Se a *expertise* inicial do aluno for básica, então sua *expertise* final será avançada;
- Se a *expertise* inicial for intermediária, então permanecerá a mesma ao final;
- Se a preferência principal for por texto, então a secundária será por *slides*;
- Se a preferência principal for por imagem, então a secundária será por vídeo;
- Tanto para o sexo feminino quanto masculino, se *expertise* inicial básica, então a *expertise* final será avançada;
- Tanto para o sexo masculino quanto feminino, se *expertise* inicial intermediária, então a mesma se manterá no final;

Para confirmar a viabilidade dos resultados obtidos, após realizar a “poda” de alguns atributos na análise e verificar a confiabilidade dos resultados obtidos, foram realizados mais dois testes com o algoritmo Apriori. Para um segundo teste, houve a retirada das preferências por materiais, o qual apontou para 18 padrões equivalentes ao primeiro teste entre *expertise* inicial e final, assim como suas relações com o gênero. Para um terceiro teste, houve a retirada das *expertises* iniciais e finais, o qual apontou para 20 padrões equivalentes ao primeiro teste com as preferências entre materiais primários e secundários, assim como suas relações com o gênero.

É importante ressaltar que as regras de associação identificadas foram analisadas com o intuito de verificar quais melhor se adequavam ao contexto proposto nas atividades do mundo virtual, sendo selecionadas estas descritas anteriormente. Assim, sua utilização poderia ser executada para que alunos dentro de cada padrão identificado pudessem receber sugestões de um agente NPC ou do próprio professor para que iniciasse suas atividades com uma determinada sequência de materiais.

Isso demonstra a viabilidade existente na tentativa de se analisar possíveis padrões de comportamento em grupos de usuários que interagem no MV. O resultado obtido pode permitir a separação de subgrupos de usuários conforme suas preferências por materiais dentro do MV, por meio da criação de regiões separadas com uma predominância maior dos tipos de materiais assinalados como preferenciais. Neste caso, diferentes regras e algoritmos de MDE poderiam ser testados em novos experimentos, como o agrupamento (clusterização), para enriquecer a análise dos dados.

Além disso, no caso das *expertises* inicial e final, os resultados demonstraram que no caso da *expertise* intermediária, que se manteve igual nesse grupo específico de alunos, poderia ser efetuado um planejamento pedagógico corretivo, com atividades de reforço específicas que ajudariam a aprimorar o conhecimento e incidir um nível final de *expertise* avançada. Vale ressaltar ainda, que os demais atributos existentes na tabela, assim como outros adicionais poderiam ser incluídos na análise para complementar as informações obtidas e fornecer um maior grau de detalhamento sobre cada grupo de usuários identificados.

7. Considerações Finais

O uso das TIC na educação proporcionou a inserção de novos meios de aprendizagem mediados pela tecnologia, dentre os quais estão os Mundos Virtuais. Dada a grande quantidade de dados obtidos das interações ocorridas entre os usuários desses ambientes, se faz necessário o uso de técnicas relacionadas à mineração de dados para possibilitar a identificação e interpretação referentes ao comportamento dos usuários.

Este artigo apresentou um estudo exploratório entre Mundos Virtuais e Mineração de Dados Educacionais, no qual foi verificada a possibilidade de identificar possíveis padrões de comportamento dos usuários dentro do mundo virtual. Um estudo de caso foi aplicado no OpenSim por meio de um laboratório virtual de química, no qual foi implementada a proposta e confirmada sua viabilidade por meio de interações simuladas dos usuários com dados sintéticos, que foram analisados com a MDE utilizando o algoritmo Apriori. Os resultados obtidos demonstram a viabilidade da proposta apresentada neste artigo, na qual foi possível identificar possíveis padrões de comportamento dos usuários dentro do laboratório virtual, relacionados à sua preferência por materiais e *expertise*.

É importante ressaltar que uma limitação deste trabalho está no fato de que os dados utilizados não foram coletados a partir de um experimento com alunos que utilizaram o ambiente, ou seja, o experimento foi realizado a partir da utilização de dados de interações simuladas de usuários. Considerando neste contexto que o uso da Mineração de Dados Educacionais se mostrou válida no contexto dos Mundos Virtuais, uma diversificada gama de possibilidades a ser pesquisada e aplicada se revela possível

para futuras implementações. Especificamente, a realização de testes com usuários reais, mineração de texto envolvendo o *chat* presente no mundo virtual, agregação de um número maior de atributos e tabelas para relacionar possíveis padrões a ser descobertos, dentre outras possibilidades a ser exploradas.

Agradecimentos

Agradecemos a CAPES pelo suporte com o Projeto Pró-Ensino na Saúde nº 39.

Referências Bibliográficas

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules in large databases. **Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)**, p.487–499, 1994.

BAINBRIDGE, W. S. Online Worlds: Convergence of the Real and the Virtual. **Human-Computer Interaction Series**, 2010.

DONALEK, C.; DJORGOVSKI, S. G.; DAVIDOFF, S.; et al. Immersive and collaborative data visualization using virtual reality platforms. **IEEE International Conference on Big Data (Big Data)**, p.1–6, 2014.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURASAMY, R. Advances in Knowledge Discovery and Data Mining. **AAAI Press**, 1996.

FERNÁNDEZ-GALLEGO, B.; LAMA, M.; VIDAL, J. C.; MUCIENTES, M. Learning Analytics Framework for Educational Virtual Worlds. **International Conference on Virtual and Augmented Reality in Education**, p.443–447, 2013.

GARRIDO, P.; MARTINEZ, F.; GUETL, C.; PLAZA, I. Enchancing Intelligent Pedagogical Agents in Virtual Worlds. **Proceedings of the 18th International Conference on Computers in Education**, v. 18, p.1–8, 2010.

GRAVAL, R.; IMIELINSKI, T.; SRIKANT, R. Mining Association Rules between Sets of Items in Large Databases. Proc. of the ACM SIGMOD Intl. Conf. on Management of Data, p.207–216, 1993.

KAMPFF, A. J. C.; FERREIRA, V. H.; REATEGUI, E.; DE, J. V.; LIMA. Identificação de Perfis de Evasão e Mau Desempenho para Geração de Alertas num Contexto de Educação a Distância. **Revista Latinoamericana de Tecnología Educativa**, v. 13, n. 2, p. 1–16, 2014.

NUNES, F. B.; HERPICH, F.; VOSS, G. B.; et al. Laboratório Virtual de Química: uma ferramenta de estímulo à prática de exercícios baseada no Mundo Virtual OpenSim. **Simposio Brasileiro de Informatica na Educacao**, p.1–10, 2014.

ROMERO, C.; VENTURA, S. Educational Data Mining: A Review of the State of the Art. **IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)**, v. 40, n. 6, p. 1–18, 2010.

YIN, R. K. Estudo de caso: planejamento e métodos. 3. ed. **Bookman**, 2005.