

Aplicação de Técnicas de Exploração de Dados como Apoio a Avaliação de Atividades Desenvolvidas em um Ambiente Virtual de Aprendizagem (AVA)

T. I. A. Souza¹, A. O. Nunes¹, F. H. L. Vasconcelos², W. B. Andriola³,
J. C. M. Mota¹, A. L. F. Almeida¹

¹ Programa de Pós Graduação em Engenharia de Teleinformática (PPGETI-UFC)
Campus do Pici, Bloco 725, CEP: 60.455-760 - Fortaleza - CE - Brasil

²Instituto UFC Virtual (UFC)
Campus do Pici, Bloco 901, CEP: 60.455-760 - Fortaleza - CE - Brasil

³Departamento de Fundamentos da Educação (UFC)
Benfica, 01, CEP: 60.020-110 - Fortaleza - CE - Brasil

{thiagoiachiley, albano.fisico}@gmail.com, herbert@virtual.ufc.br

w_andriola@yahoo.com, {mota, andre}@gtel.ufc.br

Abstract. *In recent years the expansion of distance education has brought the challenge of continuous improvement of its quality. In this context, this work by assessing student learning occurred in a course at the Federal University of Ceará (UFC), aims to identify the patterns of information obtained through the application of two techniques of multivariate statistics. Students were assessed during the development of a discipline in 14 activities in a Virtual Learning Environment (VLE). The results point to groups of standards between activities, and the most significant correlations that enable the identification of subject characteristics in relation to students.*

Resumo. *Nos últimos anos a expansão do ensino a distância tem trazido o desafio da contínua melhoria de sua qualidade. Neste contexto, o presente trabalho através da avaliação da aprendizagem discente ocorrida em um curso da Universidade Federal do Ceará (UFC), tem como objetivo identificar os padrões de informações obtidos através da aplicação de duas técnicas da estatística multivariada. Os alunos foram avaliados ao longo do desenvolvimento de uma disciplina em 14 atividades em um Ambiente Virtual de Aprendizagem (AVA). Os resultados apontam para padrões de agrupamentos entre as atividades, indicando as correlações mais significativas que possibilitam a identificação de características da disciplina em relação aos alunos.*

1. Introdução

Com o advento da Internet a sociedade vem sofrendo mudanças em suas várias esferas, sobretudo no campo da educação. A rede mundial de computadores com a evolução das Tecnologias de Informação e Comunicação (TIC) promoveu significativas transformações nas metodologias e propostas educacionais, permitindo a expansão de uma modalidade mais dinâmica e eficiente de ensino: a Educação a Distância (EAD).

A EAD se diferencia das demais modalidades de ensino não apenas pela distância física existente entre o aluno, o docente e a instituição educacional, mas também pelo uso de plataformas de ensino eletrônico. Tais sistemas, denominados como Ambientes Virtuais de Aprendizagem (AVAs), além de se caracterizarem por fornecerem um pacote de programas que comportam uma gama de funções de comunicação entre aluno e professor, proporcionam a este monitorar o progresso daquele através da utilização de vários recursos [Green 2006]. Nesta perspectiva, os AVAs permitem que sejam agrupados em um espaço virtual, recursos definidos em meios eletrônicos como: fórum, wiki, bate-papo, conferências, envios de mensagens, material de leitura, banco de questões e outras tecnologias que colaboram para o processo de ensino e aprendizagem [Rosemann 2014]. Tais recursos comportam uma significativa quantidade de informações, que no contexto da EAD podem ser utilizadas na avaliação do processo de aprendizagem discente.

O ato de avaliar é uma prática imprescindível presente tanto no sistema tradicional de educação, caracterizado pela presença do aluno em sala de aula, como na EAD. Dada essa importância, se faz necessário atentar para as diferenças existentes entre os dois sistemas. [Kuo 2014] realça essas diferenças apontando que no ensino tradicional a interação dos alunos com os professores, colegas e conteúdos é mais limitada do que na EAD. Além disso, destaca ainda que diferente da modalidade tradicional, na EAD o aluno tem mais liberdade para participar de seu próprio processo de aprendizagem. Corroborando com essas diferenças, [Karal 2012] enfatiza que no âmbito da EAD existe a necessidade de se aplicar além de métodos avaliativos que acompanham o sistema tradicional de educação, isto é, aqueles caracterizados pela cobrança de provas e trabalhos, métodos alternativos no intuito de aumentar a eficácia do processo avaliativo em termos do desempenho discente. Quando esses métodos são aplicados nessa perspectiva, o processo avaliativo permite que informações de desempenho do aluno sejam coletadas, fornecendo então subsídios ao docente para que decisões sejam tomadas em relação, às ações didáticas futuras, seleções de conteúdos, materiais de apoio, recursos tecnológicos e propostas de atividades [Piconez 2011]. Tais informações quando coletadas podem ser, por conseguinte, analisadas inclusive através de métodos de mineração de dados, especialmente da estatística multivariada.

Na literatura algumas pesquisas vêm trabalhando na perspectiva da utilização desses métodos de exploração de dados em um AVA. Em seu trabalho, [Gottardo 2012] utiliza técnicas de mineração de dados para a obtenção de um modelo de previsão de desempenho de estudantes, a partir de dados coletados em séries temporais durante o período de realização de um curso a distância. [Dias 2008] também utiliza esses métodos de exploração no intuito de descobrir informações relevantes sobre o perfil do aluno em relação ao uso dos AVAs. Na mesma linha, [Santos 2012] propõe a aplicação de métodos de mineração na predição do desempenho discente, visando identificar os possíveis indicadores de evasão na EAD.

Neste contexto, este artigo é motivado pelo fato dos AVAs gerarem grandes quantidades de informações relativas à performance do aluno, que no contexto da EAD, podem ser utilizadas para avaliar o processo de aprendizagem. Assim, é proposto neste trabalho a aplicação de duas técnicas da estatística multivariada, Análise de Componentes Principais (ACP) e Análise de Agrupamento (AA), no intuito de identificar e agrupar as atividades mais relevantes no processo de aprendizagem dos alunos. A análise busca encon-

trar padrões de comportamento nos dados, gerando ainda agrupamentos que possibilitem compreender acerca de relações não tão claras e contribuir para a avaliação discente e melhoria de seu processo de aprendizagem na EAD.

Este artigo está estruturado em 6 seções. Na seção 2, será discutido o delineamento da avaliação no processo de ensino-aprendizagem na perspectiva discente. Na seção 3, apresentam-se as descrições das duas ferramentas estatísticas utilizadas, ACP e AA. Na seção 4, descrevem-se os procedimentos metodológicos de investigação e coleta de dados desta pesquisa. Na seção 5, é apresentada a análise dos resultados obtidos com a aplicação de ambos os métodos multivariados. Por fim, na seção 6 são discutidas as considerações finais e as perspectivas de continuidade do estudo.

2. Avaliação no Processo de Ensino/Aprendizagem Discente em AVA

A utilização das TIC na educação tem sido fortemente fomentada durante a última década com a adaptação generalizada dos AVAs, que cada vez mais se destacam nesse cenário por proporcionarem um espaço de trabalho personalizado, podendo o docente projetar e instanciar atividades de aprendizagens individuais e colaborativas [Hoyos 2013]. Desta maneira, por apresentarem um conjunto de recursos tecnológicos que permitem uma maior interação e dinâmica na transmissão dos conteúdos disciplinares, esses sistemas se mostram como potenciais ferramentas no apoio da aprendizagem e no acompanhamento do processo avaliativo do discente [Sales 2011].

Neste escopo, entender como o delineamento do processo avaliativo acontece é importante para que haja planejamento no desenvolvimento dos métodos pedagógicos e eficácia quanto a sua aplicação. [Ramos 2009] destaca que essa compreensão faz-se necessária para que os objetivos propostos nos cursos sejam alcançados, ainda que se apresente como complexo. Para [Junior 2009] o sucesso do alcance dessas metas pré-concebidas reside em uma avaliação revestida de um caráter processual focada no sujeito desse processo, o aluno. Por isso, [Dario 2009] destaca que a avaliação no processo de ensino-aprendizagem oferece para esse sujeito informações que lhe permitam conhecer suas dificuldades e corrigir seus erros.

Como sujeitos do processo avaliativo, as diferentes características dos alunos devem ser levadas em conta a fim de que seus traços cognitivos e estilos de aprendizagem sejam considerados [Kurilovas 2014]. O que na EAD essa tarefa se torna desafiadora pelo fato de não haver proximidade física entre o aluno e o professor. A saída então é observar todas as etapas do processo de aprendizagem desses sujeitos, capturando todas as informações armazenadas nos AVAs a partir das atividades desenvolvidas por eles.

Neste sentido, vários trabalhos apontam para o uso das informações armazenadas nos AVAs como úteis para se compreender as etapas do processo de aprendizagem discente. Em seu trabalho, [Dias 2008] destaca bem isso apontando para a utilidade das ferramentas disponibilizadas pelos AVAs nessas etapas. Corroborando com essa ideia, [Guércio 2014] enfatiza que com o uso desses ambientes pode-se formar um grande repositório de dados educacionais com informações sobre os alunos, professores, disciplinas, bem como outras. Neste contexto [Morais 2014] defende que após o entendimento e interpretação dos dados coletados, uma investigação sobre o perfil dos cursos, professores e alunos pode ser realizada visando a melhoria e o amadurecimento dos cursos e das disciplinas. Portanto, explorar o conjunto de informações guardadas nesses sistemas,

relacionadas ao processo de ensino-aprendizagem discente, é de fundamental relevância para que o processo avaliativo se fortaleça no contexto da EAD.

3. Estatística Multivariada na Identificação de Agrupamentos

3.1. Análise de Componentes Principais (ACP)

A ACP é um método estatístico multivariado que permitir obter, a partir de um conjunto de variáveis correlacionadas, um conjunto de vetores linearmente independentes denominados de componentes principais. No método, o primeiro componente principal explica a variação máxima dos dados, já o segundo componente principal explica o desvio máximo dos dados que não foi explicada pelo primeiro e assim por diante [Santos 2013]. Seguindo essa definição, [Eder 2014] em seu trabalho descreve bem o principal objetivo da ACP: identificar por meio da redução de dimensionalidade os modos recorrentes e independentes de variações dentro de um grande conjunto de dados ruidosos, resumindo assim sua informação essencial de modo que inferências significativas possam ser estabelecidas.

Assim, o método ACP permite expressar as novas variáveis geradas como combinações lineares das informações originais. Com efeito, considera-se os n vetores de características originais $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$, com T denotando a transposta. Cada vetor \mathbf{x}_i corresponde a uma variável, possuindo um arranjo contendo os n vetores de médias $\bar{\mathbf{X}} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n]^T$, e o arranjo \mathbf{X} possuindo uma matriz de correlação denotada por $\mathbf{R}_{n \times n}$. A matriz de correlação é simétrica, além disso, não negativa. Essa condição implica que os autovalores da matriz $\mathbf{R}_{n \times n}$, denotados por, $\lambda_1, \lambda_2 \dots \lambda_n$, são não negativos, ou seja, $\lambda_i \geq 0$, estando ainda dispostos em ordem decrescente, $\lambda_1 \geq \lambda_2 \dots \geq \lambda_n$. Assim, o i -ésimo componente principal \mathbf{y}_i é definido por

$$\mathbf{y}_i = \mathbf{e}_i^T \mathbf{X} = e_{i1}\mathbf{x}_1 + e_{i2}\mathbf{x}_2 + \dots + e_{in}\mathbf{x}_n \quad (1)$$

em que o vetor \mathbf{e}_i , representado matematicamente pelos autovetores calculados da matriz de correlação, constitui as novas variáveis e estabelece a i -ésima combinação linear, para $i = 1, 2, \dots, n$. Desta forma, surge a ideia de utilizar as combinações lineares em \mathbf{y} na perspectiva de se obter uma redução dimensional, passando da dimensão n , para uma dimensão menor k , $k \leq n$. Portanto, utiliza-se as k combinações lineares principais para a análise.

3.2. Análise de Agrupamento (AA)

AA também é uma técnica da estatística multivariada que tem como objetivo agrupar variáveis com características semelhantes em grupos distintos, em função do grau de similaridade ou dissimilaridade entre os indivíduos [Ben-Dor 1999]. A técnica segrega as variáveis alocando as mais similares em um mesmo grupo, enquanto as menos similares são alocadas em outro grupo.

Nas análises exploratórias de dados os métodos de agrupamento se dividem em técnicas hierárquicas e técnicas não-hierárquicas. Para esta pesquisa, optou-se pela primeira, fazendo uso do método de [Ward 1963]. Seu objetivo é procurar por partições que minimizem a perda associada a cada agrupamento bem como identificar e determinar os possíveis grupos. Para tanto, em um algoritmo de agrupamento a similaridade ou dissimilaridade entre as variáveis é medida em função de uma distância [Cao 2012]. Ainda

segundo o autor, a função de distância mais utilizada é a métrica de Minkowski que inclui a distância Euclidiana, sendo essa a utilizada nesta pesquisa.

Logo, considerando-se novamente os n vetores de características $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$, as distâncias euclidianas entre os vetores \mathbf{x}_i e \mathbf{x}_j é dada por:

$$d(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)]^{\frac{1}{2}} \quad (2)$$

A partir da similaridade dos dados, calculada pela medida de distância da equação acima, pode-se construir um dendrograma, isto é, um gráfico em forma de árvore no qual a escala vertical indica o nível de similaridade (ou dissimilaridade), e a escala horizontal indica as variáveis agrupadas.

4. Metodologia

4.1. Procedimentos

A realização da pesquisa se deu através da investigação de alguns parâmetros extraídos da disciplina do curso de Licenciatura Plena em Física na modalidade semipresencial da Universidade Federal do Ceará (UFC). Os dados coletados são notas extraídas de planilhas fornecidas pelos professores através do AVA desenvolvido pela própria instituição denominado SOLAR. Para o experimento considerou-se as medidas de desempenho de 70 alunos avaliados na disciplina do primeiro semestre, Física Introdutória I, ao longo de 14 atividades virtuais compostas por 8 fóruns, 1 chat e 5 portfólios. Neste sentido, o experimento buscou examinar os padrões das correlações existentes entre as atividades e conteúdos executados pelos discentes.

4.2. Descrição dos Conteúdos das Atividades Analisadas

Os conteúdos trabalhados em cada atividade na qual o discente foi avaliado versam sobre os conceitos inerentes a física básica. A Tabela 1 apresenta a agenda das atividades de fórum e portfólio, especificando os conteúdos que foram abordados em cada atividade. Entretanto, a atividade de chat por apresentar discussões livres entre os alunos quanto as temáticas abordadas ao longo da disciplina, não apresentou nenhuma especificação quanto ao conteúdo, conforme observado na tabela.

	Atividade	Conteúdo
<i>Fórum</i>	1	Grandezas Físicas
	2	Movimentos
	3	Movimentos
	4	Leis de Newton
	5	Movimento Circular Uniforme
	6	Energia Mecânica
	7	Ondas Mecânicas
	8	Dilatação
Chat	9	-
<i>Portfólio</i>	10	Lista de Exercícios - Grandezas Físicas
	11	Lista de Exercícios - Movimentos
	12	Lista de Exercícios - Movimentos
	13	Lista de Exercícios - Leis de Newton
	14	Lista de Exercícios - Movimento Circular Uniforme

Tabela 1. Conteúdos das Atividades Virtuais - Física Introdutória I.

4.3. Estrutura da Disposição dos Dados

Os dados coletados nesta pesquisa, conforme descrito na subseção 4.1, foram organizados em um arranjo matricial. Os elementos que compõem as linhas dessa matriz são os alunos, enquanto que as suas colunas são as atividades desenvolvidas pelos alunos ao longo do curso da disciplina. Então, a organização dos dados compreende-se em um arranjo matricial com dimensões 70 (alunos) X 14 (atividades virtuais).

Após organizados os dados na matriz, as simulações computacionais foram realizadas no software Matrix Laboratory - MATLAB.

5. Análise e Discussão dos Resultados

A primeira etapa da análise dos resultados consiste em examinar a aplicação da ACP. Posteriormente, a segunda etapa analisa os resultados multivariados obtidos com a aplicação da AA.

5.1. Análise Multivariada dos Resultados da ACP

5.1.1. Realização dos Procedimentos da ACP

Antes da aplicação da ACP, o conjunto de dados foi submetido a dois testes de adequação da amostra para a aplicação do método. O primeiro teste indica a proporção da variância dos dados que podem ser consideradas comuns a todas as variáveis. Essa indicação varia no intervalo de 0 à 1, em que, quanto mais próximo de 1 (unidade) melhor é o resultado, ou seja, os dados se adequam a aplicação da ACP. Já o segundo teste, verifica se as variáveis são correlacionadas entre si gerando a hipótese da matriz de correlação desses parâmetros ser a matriz identidade [Green 2011].

O resultado dos testes validaram o uso do método multivariado (Tabela 2). Com um valor de 0,6 para o teste de KMO, os dados se adequam à aplicação do método multivariado, e com um valor de significância inferior a 0,05, o teste de Bartlett rejeita a hipótese nula da matriz de correlação ser uma matriz identidade.

Após a validação dos testes, utilizou-se dois critérios clássicos para a seleção do número de componentes principais. O primeiro se baseia na observação da quantidade de informação que cada componente principal armazena, representada pelo valor percentual da variância explicada. Já o segundo fundamenta-se na retenção dos autovalores superiores à unidade, conhecido como critério de Kaiser [Green 2011]. A Tabela 2 discrimina apenas os autovalores que satisfazem a esse critério e suas respectivas variâncias explicada. Portanto, dos 14 componentes principais calculados pela ACP, selecionou-se os 5 primeiros que explicam mais de 70% da informação.

Valores dos Testes de Validação					
Componente Principal	Física Introdutória I				
	1	2	3	4	5
Autovalor	3,572	2,522	1,676	1,338	1,002
Variância (%)	25,511	18,014	11,975	9,557	7,158
Variância Acumulada (%)	25,511	43,525	55,500	65,057	72,216
KMO	0,600				
Significância de Bartlett	0,000				

Tabela 2. Testes de Validação.

5.1.2. Análise dos Padrões da ACP

A partir da retenção do número de componentes principais, buscou-se explorar as relações entre as variáveis desta pesquisa no novo espaço gerado. Tal relação é perceptível através da plotagem gráfica das variáveis originais no espaço dos dois primeiros componentes principais. A análise das relações entre os componentes principais busca fortalecer as interpretações mais consistentes, permitindo que padrões no comportamento entre as atividades avaliadas possam ser extraídos e examinados.

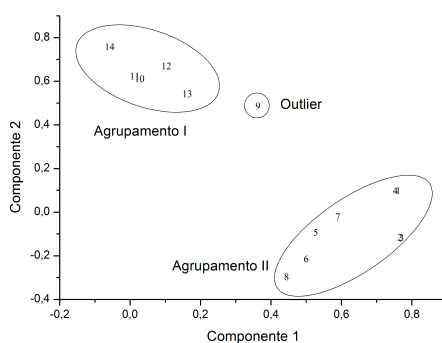


Figura 1. Agrupamento - Física Introdutória I

Verifica-se na Figura 1 a formação de dois agrupamentos e um outlier (Atividade 9 - Chat). Tais configurações nos padrões de comportamento retornados pela ACP apontam para uma tendência presente nos dados que se manifesta nos agrupamentos.

No intuito de garantir que esses padrões de agrupamentos apresentem confiabilidade, aplicou-se o coeficiente α de Cronbach na perspectiva de mensurar a consistência interna de cada grupo gerado pela ACP [Cronbach 1951]. Assim, com um $\alpha = 0,87$ para o agrupamento I, e um $\alpha = 0,73$ para o grupo II, destaca-se a alta consistência interna dos dados, garantindo a uniformidade dos agrupamentos formados.

Analisando o primeiro grupo, as atividades do Agrupamento I, têm em comum o fato de serem as atividades de portfólio, havendo destaque para os altos fatores de carregamento para o segundo componente principal. Logo, pode-se interpretar o componente 2 como aquele relacionado às atividades postadas nos portfólios. Isto indica o portfólio como um instrumento facilitador do processo de ensino e aprendizagem para a formação discente no contexto da EAD. Ou seja, o uso do portfólio oferece uma nova metodologia que descentraliza o antigo modelo pedagógico, como a aplicação tradicional de provas, da figura do professor e foca no aluno lhe oferecendo motivação, responsabilidade e exercício da autonomia.

O Agrupamento II reúne as atividades de fórum, estando as correlações associadas fortemente com o primeiro componente principal. Portanto, infere-se que o componente 1 é aquele que reflete o forte grau de participação dos discentes nas atividades realizadas nos fóruns. Nesta perspectiva, a análise destaca a utilização do fórum como um recurso fundamental de comunicação entre os alunos, seus colegas e os professores. Desta forma, além de proporcionar a utilização dos suportes tecnológicos disponíveis como canal de comunicação, o fórum permite que se estabeleça conexões entre os participantes, reunindo

os alunos em torno de um mesmo objetivo de aprendizagem.

Por fim o outlier (Atividade 9 - Chat), encontra-se isolado dos dois agrupamentos. Sendo o chat também uma ferramenta de comunicação voltada para a discussão de temas relacionados à aprendizagem, tal isolamento se justifica quando o comparamos com os dois grupos. Primeiro, quando comparado com o agrupamento das atividades de fórum, infere-se que o chat é vantajoso por permitir uma comunicação síncrona (em tempo real), proporcionando discussões interativas entre assuntos distintos. Segundo, quando comparado com o agrupamento das atividades de portfólio, infere-se que enquanto no chat permite avaliar a aprendizagem através da interação nas discussões inerentes aos conteúdos entre os alunos, nos portfólios se observa a evolução do aluno sob a perspectiva do amadurecimento do conhecimento através dos materiais produzidos e postados em todas as fases do curso.

5.2. Análise Multivariada dos Resultados da AA

Os resultados da aplicação da AA, discutidos nesta seção, tem a finalidade de complementar a análise iniciada pela ACP. Desta forma, através do método de Ward, buscou-se identificar e determinar os possíveis agrupamentos não encontrados pela ACP.

Conforme discutido na seção 3, a distância Euclidiana foi a medida utilizada para calcular as distâncias entre as variáveis e assim agrupa-las. Tais agrupamentos são vislumbrados na Figura 2 através do dendrograma.

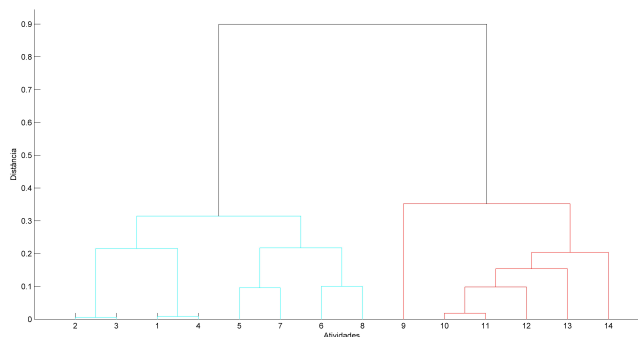


Figura 2. Dendrograma - Física Introdutória I

A análise através do dendrograma permite identificar dois agrupamentos. Observa-se que o primeiro grupo abrange as atividades de fórum, corroborando com o agrupamento II da ACP. Já o segundo grupo engloba as atividades de portfólio, incluindo o chat. Entretanto, como observado no eixo das distâncias a atividade de chat apresenta uma medida de distância relativamente superior às demais variáveis de seu grupo podendo, portanto, ser considerada como um outlier. Tal inferência também concorda com o resultado da ACP, que agrupa as atividades de portfólio em um segundo agrupamento e a atividade de chat como outlier.

6. Conclusões e Trabalhos Futuros

Apresentou-se nesta pesquisa um estudo qualitativo a partir de uma análise quantitativa acerca do processo da aprendizagem discente no curso de Licenciatura Plena em Física

na modalidade semipresencial da Universidade Federal do Ceará. Para tanto, os discentes matriculados na disciplina do primeiro semestre do curso, Física Introdutória I, foram avaliados ao longo de 14 atividades.

As análises discutidas neste trabalho podem fornecer aos professores do curso subsídios para a melhoria na qualidade do seu ensino. Os resultados apontam para o potencial uso da atividade de fórum como instrumento a ser considerado no decorrer do processo da aprendizagem discente.

Os resultados dos testes estatísticos de validação multivariados mostraram-se significativos à base de dados educacional usada nesta pesquisa. Tais análises não apenas proporcionaram resultados estatísticos, mas também apontaram e indicaram possíveis ações práticas à serem tomadas pelos docentes quanto ao uso de determinadas ferramentas virtuais de comunicação afim de potencializar as interações dos alunos com os conteúdos abordados e buscar assim formas alternativas de avaliar a aprendizagem discente através da participação nessas atividades.

A aplicação dos dois métodos da estatística multivariada, ACP e AA, mostraram-se como potenciais ferramentas na análise e exploração de dados relacionados a avaliação do processo da aprendizagem discente. Assim, padrões no comportamento das informações analisadas foram vislumbrados e percepções acerca da aprendizagem discente tornaram-se mais claras.

Como trabalho futuro sugere-se uma análise mais minuciosa dos agrupamentos gerados por cada método multivariado. Há ainda a possibilidade de se utilizar outras técnicas multivariadas de identificação e classificação de grupos bem como estabelecer uma análise na perspectiva do professor.

Referências

- Ben-Dor, A.; Shamir, R. (1999). Clustering gene expression patterns. *Journal of Computational Biology*, 6:281–297.
- Cao, F.; Liang, J. (2012). A dissimilarity measure for the k-modes clustering algorithm. *Knowledge-Based Systems*, 26:120–127.
- Cronbach, L. (1951). Coeficient alpha and the internal structure of tests. *Psychometrika*, 16:297–334.
- Dario, D.; Rocha, H. V. (2009). Aplicação de técnicas de visualização de informação em ferramentas para apoio à avaliação formativa em sistemas de ensino a distância. *Anais do XX Simpósio Brasileiro de Informática na Educação*.
- Dias, M.; Filho, L. (2008). Aplicação de técnicas de mineração de dados no processo de aprendizagem na educação a distância. *Anais do XIX Simpósio Brasileiro de Informática na Educação*, pages 104–114.
- Eder, B.; Foley, K. (2014). Incorporating principal component analysis into air quality model evaluation. *Atmospheric Environment*, 82:307–315.
- Gottardo, E.; Kaestner, C. (2012). Previsão de desempenho de estudantes em cursos ead utilizando mineração de dados: uma estratégia baseada em séries temporais. *Anais do XXIII Simpósio Brasileiro de Informática na Educação*, pages 26–30.

- Green, S. M.; Weaver M.; Voegeli, D. (2006). The development and evaluation of the use of a virtual learning environment (blackboard 5) to support the learning of pre-qualifying nursing students undertaking a human anatomy and physiology module. *Nurse Education Today*, 26:388–395.
- Green, P. E. (2011). *Multivariate Data Analysis*. Cengage Learning.
- Guércio, H.; Marques, P. (2014). Análise do desempenho estudantil na educação a distância aplicando técnicas de mineração de dados. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, pages 641–650.
- Hoyos, C.; Lorenzo, M. (2013). Glue!: An architecture for the integration of external tools in virtual learning environments. *Computers & Education*, 60:122–137.
- Junior, L., D. D. (2009). Discutindo os caminhos da avaliação para uma melhor aprendizagem on-line. *Anais do XX Simpósio Brasileiro de Informática na Educação*.
- Karal, H.; Cebi, A. (2012). Views on modular assessment and evaluation process in distance education. *Procedia Social and Behavioral Sciences*, 46:2073–2077.
- Kuo, Y.; Walker, A. (2014). Interaction, internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *Internet and Higher Education*, 20:35–50.
- Kurilovas, E.; Kubilinskiene, S. (2014). Web 3.0 - based personalisation of learning objects in virtual learning environments. *Computers in Human Behavior*, 30:654–662.
- Morais, A.; Araujo, J. (2014). Perfil do docente em educação a distância e suas perspectivas frente aos processos avaliativos. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, pages 242–251.
- Piconez, S. ; Nakashima, R. H. (2011). Avaliação em ambientes virtuais de aprendizagem: articulação dialética de suas dimensões. *Anais do XXII Simpósio Brasileiro de Informática na Educação*, pages 2225–2234.
- Ramos, J.; Gomes, A. (2009). Processo de elicitação de requisitos para sistemas de avaliação em ambientes virtuais de aprendizagem. *Anais do XX Simpósio Brasileiro de Informática na Educação*.
- Rosemann, D.; Raabe, A. (2014). Personalização de conteúdo e avaliação multicritério em ambiente virtual de aprendizagem de código aberto. *Anais do VII WAvalia*.
- Sales, G., L. B. G. (2011). Um indicador de aprendizagem não-linear para ead online fundamentado no modelo de avaliação learning vectors (lv). *Revista Iberoamericana de Evaluación Educativa*, 4:155–180.
- Santos, P.; Cremona, C. (2013). Multivariate statistical analysis for early damage detection. *Engineering Structures*, 56:273–285.
- Santos, H.; Camargo, F. (2012). Predizendo o sucesso de estudantes através do uso avaliações formativas em avas. *Anais do XXIII Simpósio Brasileiro de Informática na Educação*.
- Ward, J. (1963). Hierarchical grouping of optimize an objective function. *Journal of the American Statistical Association*, 58:236–244.