

Evolução das posturas corporais de estudantes em apresentações de seminários ao longo do semestre: uma análise utilizando dados multimodais e técnicas de clusterização

**Felipe Vieira¹, Cristian Cechinel¹, Roberto Munoz², Robson Lemos¹,
Erick Merino², Rodolfo Villarroel³**

¹Programa de Pós-Graduação em Tecnologias da Informação e Comunicação,
Universidade Federal de Santa Catarina, Araranguá, Brasil

²Escuela de Ingeniería Civil Informática,
Universidad de Valparaíso, Valparaíso, Chile

³Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

felipe.roque@posgrad.ufsc.br, {cristian,robson}@ufsc.br, roberto.munoz@uv.cl,

erick.merino@postgrado.uv.cl, rodolfo.villarroel@pucv.cl

Abstract. *The present paper describes a clustering based approach to identify the main corporal patterns in students oral presentations during a given course. Data from 43 students presentations was collected through the use of Microsoft Kinect. The 16 collected features were used as input information in the clusterization process allowing the identification of three main profiles of presenters: passive, active, and semi-active. An analysis of the evolution of these profiles during the semester points out a decrease in the percentage of the passive profile throughout the course, and an increase in the percentage of the semi-active profile. These different profiles will be integrated into the system that collects the postures information in order to allow the automated classification of the presenters in real time.*

Resumo. *O presente trabalho apresenta uma abordagem baseada em clusterização para identificar os principais padrões corporais em apresentações orais de estudantes em uma disciplina. Utilizando o Microsoft Kinect, foram coletados dados de 43 apresentações de estudantes realizadas em três momentos distintos ao longo do semestre. As 16 características coletadas pelo sistema desenvolvido foram utilizadas como entrada para a clusterização que permitiu identificar três perfis principais de apresentadores: passivos, ativos e semi-ativos. Uma análise sobre a evolução desses perfis aponta que houve uma diminuição do percentual do perfil passivo ao longo do semestre e um aumento do percentual do perfil semi-ativo. Esses tipos de perfis serão integrados ao sistema de coleta de posturas para a futura classificação automática dos apresentadores em tempo real.*

1. Introdução

Bons comunicadores demonstram suas habilidades de apresentação com características verbais e não verbais, como a linguagem corporal, o contato visual com a audiência, ou

também pelo espaço que ocupam no palco [York 2013]. As posturas de um aprendiz podem fornecer informações importantes sobre seu estado interior [Ochoa 2017]. Avaliar ambientes complexos de aprendizagem pode ajudar os professores a oferecer um *feedback* aos seus alunos assim como também auxiliar o mesmo a compreender melhor o ambiente de ensino-aprendizagem em que está inserido.

As instituições de ensino têm trabalhado de forma a proporcionar aos estudantes o desenvolvimento de suas habilidades de comunicação [Ochoa et al. 2018]. Isso pode ser particularmente observado na grande disseminação do uso de apresentações orais em formato de seminários como parte do processo avaliativo dos cursos lecionados.

A área de *Multimodal Learning Analytics* (MLA) [Leong et al. 2015] trabalha diretamente com a captura, análise e utilização de informações combinadas e normalmente não convencionais, ou que não são regularmente obtidas por meio dos Ambientes Virtuais de Aprendizagem e Sistemas de Informação Acadêmicos. Dados como áudio, vídeo, posturas e medidas corporais (batimentos cardíacos, pulsação, temperatura, etc.) são exemplos de informações multimodais que vem sendo gradativamente incorporadas aos sistemas voltados para a aprendizagem.

Problemas com dados oriundos de diversas fontes e programas e que envolvem o reconhecimento facial e de fala normalmente possuem resolução complexa, que muitas vezes podem ser resolvidos por meio de técnicas de aprendizado de máquina [Duda et al. 1995]. O presente trabalho apresenta uma abordagem para a identificação de padrões em apresentações de estudantes utilizando dados multimodais provenientes da fala e de informações corporais dos estudantes coletadas por meio do Microsoft Kinect. A abordagem consiste na aplicação e avaliação de técnicas de clusterização (*Silhouettes* e *K-means*) para tentar encontrar categorias e perfis distintos desses apresentadores. O objetivo final é o desenvolvimento de um sistema capaz de reconhecer automaticamente os perfis dos estudantes apresentadores durante a realização de uma apresentação oral, de maneira a permitir um *feedback* imediato sobre as apresentações realizadas. Para o presente trabalho, a seguinte pergunta de pesquisa é proposta: *É possível utilizar dados multimodais e técnicas de aprendizado de máquina para acompanhar a evolução dos padrões corporais dos estudantes em apresentações orais ao longo do semestre?*

O restante do artigo está organizado da seguinte forma. A Seção 2 são apresentados os trabalhos relacionados com a identificação de padrões posturais. Já na Seção 3 é apresentada uma visão geral da metodologia e das etapas realizadas para obtenção dos resultados. Os dados coletados são descritos na Seção 4 e na Seção 5 são apresentados os resultados encontrados. A Seção 6 apresenta as considerações finais e propostas de trabalhos futuros.

2. Trabalhos Relacionados

Alguns trabalhos já foram propostos para identificar padrões posturais em diversos contextos. Essa seção apresenta alguns desses trabalhos e as principais técnicas utilizadas. [Echeverría et al. 2014], utilizou uma abordagem de vídeo conjuntamente com o Kinect para extrair os dados das apresentações individuais de estudantes. A qualidade das apresentações orais foram avaliadas. Após pre-processar e selecionar as principais características, os autores utilizaram o classificador *Weighted Logistic Regression* para avaliar os dados em separado. Também utilizando o MS Kinect, o trabalho de [Le et al. 2013]

explora a capacidade do uso de informação de *skeleton* que o Kinect oferece para reconhecer posturas humanas. O dispositivo desenvolvido permite reconhecer corretamente as posturas de deitar, sentar, levantar e se curvar por meio de um classificador *Support Vector Machines* (SVM).

[Chen et al. 2015] utilizou o Kinect em um estudo sobre o desenvolvimento de um modelo automático de pontuação de apresentações em público interpretando dados multimodais. Os autores utilizaram algoritmos de aprendizado de máquina (SVM, *Random Forests* e *glmnet*) para realizar as pontuações com base em características das posturas, olhos, expressão facial, traços de movimento, entre outras. [Schneider and Blikstein 2015] utilizou técnicas de clusterização em dados coletados pelo Kinect para avaliar estratégias de estudantes interagindo com uma TUI *Tangible User Interface*. De acordo com os autores, os estudantes apresentaram três tipos principais de posições: ativa, semi-ativa, e passiva. Os autores também afirmam que o tempo gasto pelos estudantes no estado ativo está diretamente relacionado com o ganho de aprendizagem, e que o tempo gasto no estado passivo está diretamente relacionado com o baixo ganho de aprendizagem.

[Ochoa et al. 2018] utiliza técnicas de aprendizado de máquina não supervisionado para analisar dados multimodais (posturas, áudio, olhar, entre outros) e desenvolver um sistema de *feedback* automático para estudantes no contexto de apresentações orais. Por último, [Junokas et al. 2018] utiliza uma abordagem supervisionada (modelo hidden-Markov hierárquico) para o reconhecimento automático de gestos e com informações de entrada relacionadas a posições de *skeleton*, características e parâmetros internos de movimento.

3. Metodologia

Os passos seguidos na metodologia do trabalho são descritos na Figura 1. A primeira etapa consiste na extração das características utilizadas no problema. A segunda etapa consiste na utilização do *K-means* para gerar os grupos com base nas características extraídas. A terceira etapa é a análise, interpretação e validação da consistência dos clusters gerados por meio do método *Silhouette*. A quarta etapa é a avaliação da evolução dos clusters ao longo do semestre. A quinta etapa consiste na utilização das classes geradas para um experimento inicial de classificação automática dos estudantes de acordo com o seu perfil.

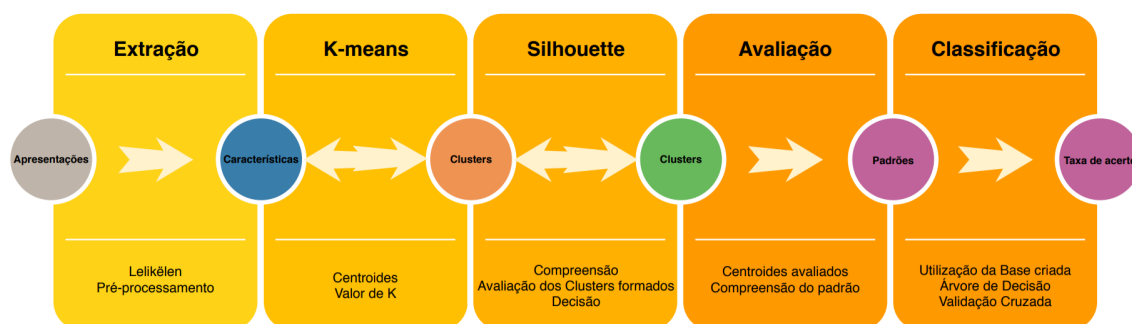


Figura 1. Metodologia utilizada no trabalho.

3.1. Extração

A etapa de extração consistiu em algumas etapas até a montagem da base de dados propriamente dita. Cada atributo do problema teve seus intervalos de ocorrência em cada apresentação computados. Assim, um atributo pode ocorrer em diferentes momentos da apresentação, de acordo com o que o estudante faz durante a mesma. O software Lelikëlen foi responsável por coletar e gerar a base de dados *crua* com esses intervalos.

Na etapa de pré-processamento foi realizada a soma total de cada intervalo para cada uma das 16 características coletadas. Os intervalos em questão eram uma parte do total da apresentação. Dessa forma, o máximo possível para uma determinada observação de um determinado atributo é 1, equivalente a 100% do tempo da apresentação. Para cada observação da base de dados foi computada a soma de seus dados (características) analisados em uma apresentação resultando em uma matriz onde as colunas representam os atributos (características posturais) e as linhas representam as observações.

3.2. K-means

O k-means[Lloyd 1982] é um algoritmo iterativo de particionamento de dados que atribui i observações para exatamente um dos k clusters definidos pelos centroides. O valor de k é definido antes do algoritmo iniciar e o agrupamento é alcançado minimizando a soma das distâncias dos quadrados entre os dados e os centroides dos clusters. Neste trabalho, o algoritmo *K-means* foi utilizado para particionar os dados dos estudantes em k clusters.

3.3. Silhouettes

Para explorar os dados e analisar os grupos criados pelo K-means, o algoritmo de análise visual Silhouette [Rousseeuw 1987] foi usado. Esse método é útil para selecionar o melhor número de clusters pré-definidos, por exemplo. Ele também pode ser usado para trocar uma observação que possui valor negativo da silhouette para o seu vizinho, melhorando assim os resultados da análise dos grupos [Rousseeuw 1987]. O valor da Silhouette é definido pela Equação 1, como segue:

$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]} \quad (1)$$

Onde $a(i)$ é a distância média entre um exemplo i e todos os outros dados dentro do mesmo cluster. Isso implica que $a(i)$ é o quanto bem determinado objeto i está dentro de seu cluster. $b(i)$ é a menor distância média de i para todos os pontos em qualquer outro cluster de quais i não faz parte. Logo, quanto mais perto de 1 o valor de $s(i)$, mais pertencente ao cluster atribuído i é. Analogamente, valores negativos representam que este exemplo não pertence ao cluster. Valores muito próximos de 0 estão no limite entre os clusters, podendo então variar entre os dois.

3.4. Avaliação

A avaliação dos resultados obtidos é feita em diferentes momentos e mais de uma vez a cada rodada dos algoritmos. O motivo principal para essa avaliação ocorrer em diferentes momentos é que tanto o algoritmo *K-means* quanto o algoritmo de Silhouettes retornarem resultados em suas próprias etapas que impactam no resultado final como um todo. Exemplificando, a atribuição dos clusters do *K-means* é avaliada pelo algoritmo de Silhouettes

para verificar se o agrupamento foi bom ou não. Contudo esse resultado não pode ser avaliado sozinho, devendo os centroides ser avaliados de modo a encontrar e verificar se determinado padrão esperado está ou não acontecendo. Dessa forma, essas etapas se repetem algumas vezes durante o processo de descoberta dos padrões.

3.5. Classificação

A classificação dos padrões encontrados tem o objetivo de avaliar a base de dados com os rótulos. Para este trabalho foi utilizado o classificador de Árvore de Decisão por sua facilidade de interpretação. Com esse resultado, o *feedback* ao aluno pode ser dado facilmente uma vez que o classificador seja integrado ao sistema de coleta das informações corporais.

4. Extração de características e base de dados

4.1. Dados

Os dados usados nesse trabalho foram organizados/extraídos com o software Lelikëlen. O software pode armazenar, detectar e ainda visualizar posturas corporais [Munoz et al. 2018]. Para a captura dos dados foi usado o Microsoft Kinect. Para a classificação das posturas corporais, o software utiliza o algoritmo ADAboost [Freund and Schapire 1995].

As características utilizadas no trabalho são apresentadas na Tabela 1. Como mencionado, esses dados são capturas de ações dos aprendizes. Esses dados em forma bruta eram intervalos de tempos das ações que eram realizadas. Um estudante, por exemplo, pode ficar um tempo determinado com os braços cruzados, após um período, pode deixar os braços para baixo e começar a falar e por fim, retornar a posição de braços cruzados. Todos esses tempos em ações são classificados e computados pelo software Lelikëlen.

4.2. Base de Dados

O contexto de aprendizagem para esse experimento foram apresentações orais realizadas por estudantes de graduação. Foi solicitado aos estudantes a realização dessas apresentações de projetos com tópicos previamente definidos [Munoz et al. 2018]. Essas apresentações foram realizadas em duplas (exceto por uma apresentação que foi realizada por apenas um estudante), que foram sorteadas em uma turma com aproximadamente 90 alunos. Ao longo do semestre, foram realizadas 3 apresentações. Para a primeira, segunda e terceira apresentações foram compostas por 40, 22 e 23 observações respectivamente, totalizando 85 observações.

A base de dados é composta por 16 características apresentadas na Tabela 1 e por 85 observações. Geralmente, uma observação E , é representada por uma coleção de valores de atributos (características) $(x_1; x_2; \dots; x_n)$, onde x_i é um valor de um atributo X_i . Para este problema, um exemplo é um estudante em uma apresentação. Como não se tem as classes para realizar a classificação dos exemplos, este é um problema que deve ser abordado por meio do aprendizado não supervisionado.

Tabela 1. Descrição de cada característica da base de dados inicial

N	Características	Descrição
1	Fala	Foi detectada a voz do apresentador.
2	Braços Cruzados	O apresentador cruzou os dois braços.
3	Mão na Face	O apresentador tem uma mão no queixo.
4	Mão na Cabeça	O apresentador tem uma mão na nuca.
5	Mãos no Quadril	O apresentador tem as mãos na cintura.
6	Mãos Baixas	O apresentador tem as mãos para baixo.
7	Uma Mão	O apresentador está explicando com uma mão para baixo e a outra dobrada em posição explicativa.
8	Mãos Abertas	O apresentador está explicando com as duas mãos (as duas mãos com os braços dobrados).
9	Apontando	O apresentador está apontando com uma mãos (braço estendido horizontalmente)).
10	Olhando o Público	O apresentador está olhando o público.
11	Público	A distância entre o ponto médio(externo) dos que apresentam é maior ou igual a 85cm.
12	Social	A distância entre o ponto médio (externo) dos que apresentam é maior ou igual a 65 e 85cm.
13	Pessoal	A distância entre o ponto médio (externo) dos que apresentam é maior ou igual a 45 e 65 cm.
14	Íntimo	A distância entre o ponto médio (externo) dos que apresentam é menor que 45cm.
15	Para Baixo	A inclinação do apresentador é maior que 0,333, com -1 inclinada para trás e 1 inclinada para a frente.
16	Posição Ereta	A inclinação do apresentador está entre -0,333 e 0,333, com -1 inclinada para trás e 1 inclinada para a frente.

5. Resultados - Avaliação do contexto de aprendizagem

Com a intenção de avaliar o contexto, de modo que padrões de apresentação sejam encontrados, alguns algoritmos foram utilizados para explorar os dados. O algoritmo *K-means* agrupa exemplos similares ao mesmo conjunto, de acordo com suas características. Dessa forma, pode-se analisar os valores dos *centroides*. Mesmo assim, em muitos casos é difícil e diversas vezes não se sabe quantos grupos o problema possui. O algoritmo de *Silhouettes* auxilia nesse sentido, ajudando visualmente na identificação de casos onde o *K-means* está posicionando de forma errada algum exemplo.

Dessa forma, foram realizados testes com valores de média de *Silhouettes* e foi identificado que o melhor número de $k = 3$ (para o *K-means*). Para $k = 2$, o valor médio foi melhor do que para 3, entretanto algumas informações necessárias se perdiam com apenas dois grupos. Para valores de k maiores que 3, os resultados se mostraram insatisfatórios, com valores médios de *Silhouettes* mais baixos. A Figura 2 apresenta os resultados de *Silhouettes* para $k = 3$ em cada uma das apresentações ao longo do semestre. Analisando as Figuras é possível visualizar que o melhor resultado para as *Silhouettes* está na Figura 2(c), onde os valores se aproximam mais de 1. Isso pode indicar que os estudantes foram ao longo do semestre "padronizando" seu comportamento

de apresentação a medida que apresentavam e viam apresentações.

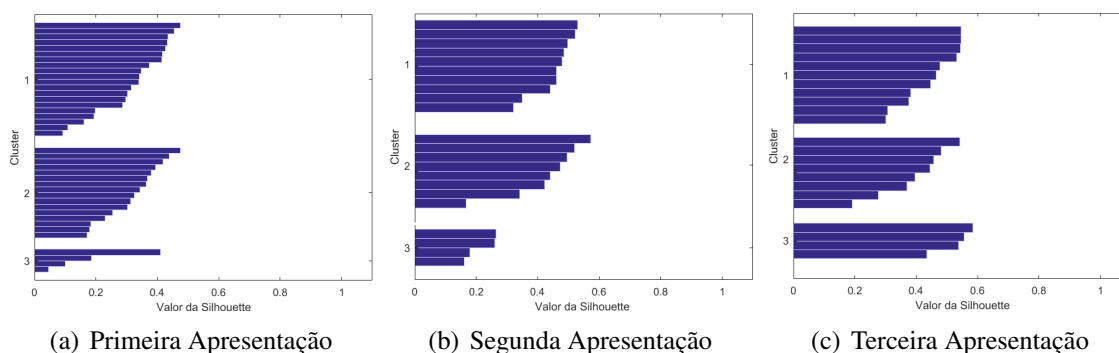


Figura 2. Gráficos das Silhouettes para diferentes momentos do semestre.

Ainda para a avaliação dos grupos formados, foram observados os valores dos *centroides* de cada atributo da base de dados formada. Dessa forma, pode-se encontrar algum padrão na formação dos grupos de acordo com os valores dos *centroides*. Como existem três grupos diferentes, foram analisados os centroides de cada um dos grupos para cada variável do problema. A Figura 3 apresenta alguns valores de *centroides* para as variáveis de maior influência nos clusters.

A partir dos valores dos centroides foram identificados três comportamentos distintos em todas as três apresentações. Estes comportamentos foram rotulados como: Ativo, Semi-Ativo e Passivo, utilizando-se como base os comportamentos já identificados por [Schneider and Blikstein 2015] no contexto da utilização de TUI. Estes grupos tem relação com os valores das variáveis de análise que estão apresentadas na Figura 3 e que podem evidenciar os padrões dos mesmos (a figura apresenta apenas 4 das 16 características analisadas). Por exemplo, os apresentadores do grupo ativo possuem o atributo *Olhando o Público* sempre baixo em todas as apresentações. Este também é o grupo onde o atributo de *Fala* se destaca mais. Já o atributo de *Braços Cruzados* tem um valor baixo para este grupo (atributo não incluído na figura).

O grupo Semi-ativo tem comportamento parecido com os outros dois grupos. Este grupo, a partir da apresentação 2, apresenta o valor da variável *Apontando* mais próximo ao do grupo Ativo (atributo que está relacionado com a intenção de explicar algo para o público). Este grupo olha bastante para o público (*Olhando para o Público*) também, similar ao grupo Passivo e diferente do grupo Ativo. A característica de *Fala* deste grupo é maior que a do grupo Passivo e menor do que o Ativo em algumas apresentações. O terceiro grupo, chamado Passivo, tem a característica de olhar para o público muito similar com o grupo Semi-ativo. A postura de mãos baixas é a maior entre todos os três (atributo não incluído na figura). Este grupo também é o que menos aponta (*Apontando*). Esse grupo fica mais tempo com uma mão para baixo e outra dobrada em posição explicativa (*Uma Mão*) do que o grupo Semi-ativo na apresentação 1, e menos do que o grupo Ativo. Por fim, este grupo é o que menos fala entre todos (falando menos do que a metade do que o grupo Ativo em algumas apresentações).

É possível observar que os padrões dos grupos acontecem desde a primeira apresentação. Contudo, pode-se observar que na última apresentação os grupos ficaram

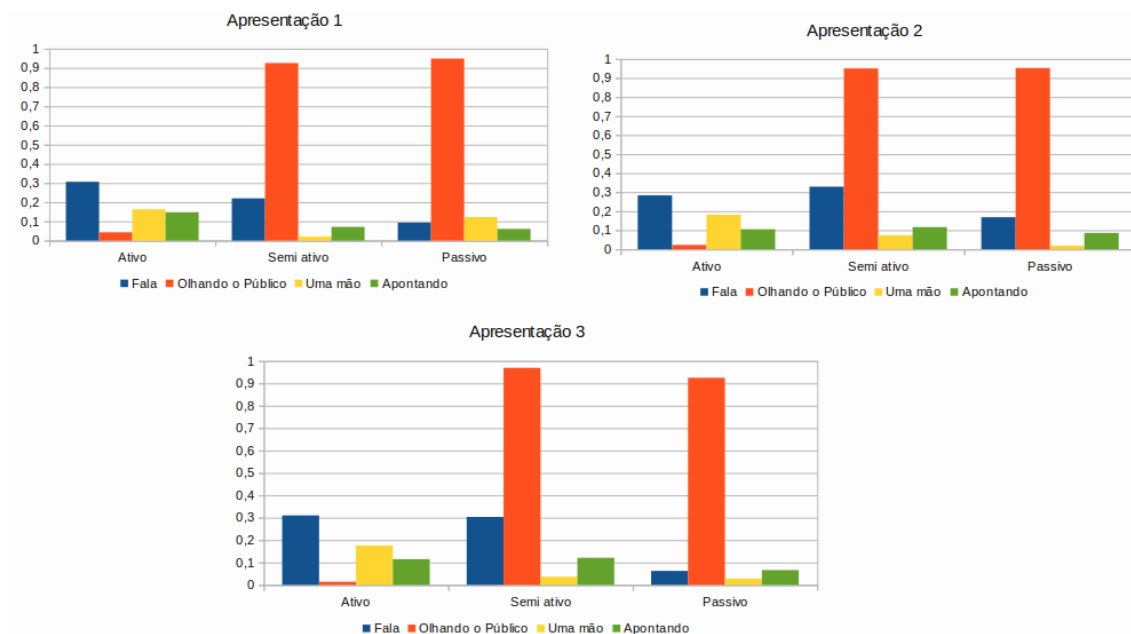


Figura 3. Valores dos centroides para cada característica de acordo com os grupos

ainda mais definidos. O grupo Passivo, por exemplo, se diferenciou mais do grupo Semi-ativo nessa última apresentação. Os valores de *Fala* e *Uma Mão* vão diminuindo gradativamente até a terceira apresentação. Vale ressaltar que os outros atributos também compõem papel importante para a formação dos clusters e estão omitidos aqui por questões de espaço.

Como o número de observações nas apresentações ao longo do semestre é diferente, foi realizada uma análise dos percentuais gerais de estudantes pertencentes a cada grupo em cada apresentação. Essa análise tem como objetivo avaliar a evolução das posturas ao longo do semestre dentro da disciplina. A Figura 4 apresenta os dados dessas proporções. Inicialmente, a proporção de estudantes com características ativas era de 50% do total na apresentação 1. Essa proporção diminuiu na segunda apresentação (45, 45%) e subiu para 47, 83% na última apresentação. Os semi-ativos eram 10% na primeira apresentação, tendo subido para 18, 18% na segunda e representando 17, 39% na última. Pode-se afirmar que houve um aumento proporcional considerável nessa classe entre a primeira e a terceira apresentação. O grupo passivo obteve 40% do total na primeira apresentação, 36, 36% do total da segunda e 34, 78% do total da terceira apresentação. É possível observar que a proporção desse grupo foi gradativamente diminuindo ao longo das apresentações. Isso pode ser um indicativo de uma certa melhoria nas posturas dos estudantes ao longo das apresentações, ainda que esses resultados iniciais necessitem de uma análise mais aprofundada.

Os padrões (rótulos) encontrados pela análise realizada foram utilizados para a geração de um modelo de classificação capaz de identificar os estudantes com base em suas posturas. A base contendo as 85 apresentações foi rotulada com as classes Ativo, Semi-ativo e Passivo. Utilizando o algoritmo J48 (para geração de árvores de decisão) e a técnica de validação cruzada de 10 partições foi possível obter uma taxa geral de acurácia

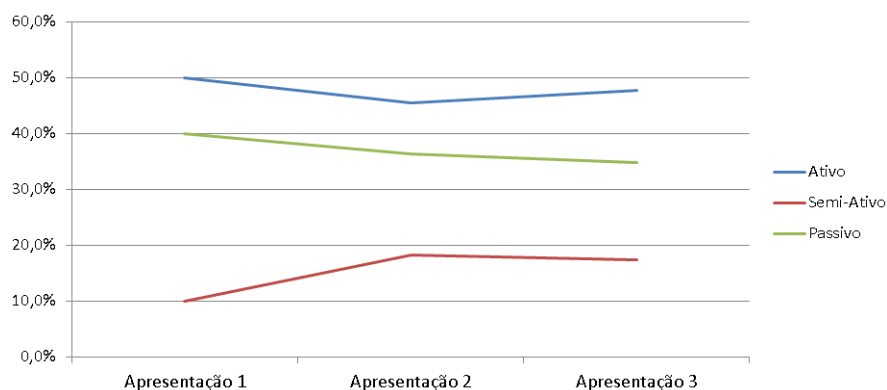


Figura 4. Evolução percentual dos grupos ao longo do semestre

de 94,1%. Em uma próxima etapa do trabalho, modelos de classificação serão integrados a ferramenta desenvolvida para a identificação automática do perfil de cada apresentador após a realização de sua apresentação.

6. Considerações finais

O *Multimodal Learning Analytics* pode auxiliar na avaliação de ambientes de aprendizado complexos usando dados de diversas fontes. No presente estudo, uma análise inicial mostra que os estudantes evoluíram de características passivas para características semi-ativas ao longo do semestre. Ao passo que a migração ocorreu do grupo passivo para os grupos ativo e semi-ativo. 47,83% do total de estudantes na apresentação 3 ficaram no grupo Ativo, 17,39% ficaram no grupo Semi-ativo e 34,78% terminaram no grupo Passivo. Ressalta-se que o número de observações de 85 para análise foi pequeno, o que pode ter prejudicado a classificação do Lelikelen e ainda a clusterização dos padrões. Por outro lado, a classificação dos padrões encontrados apresenta uma taxa de acerto alta, com 94,1%, fato que pode indicar que os dados contém certo padrão.

Como trabalhos futuros espera-se fazer novas análises em diferentes grupos de estudantes. Esses novos estudantes seriam de diferentes áreas de atuação, como engenharias, saúde e ciências humanas, por exemplo. Essa distinção poderia evidenciar diferentes comportamentos entre as áreas de conhecimento. Ou ainda poderia revelar comportamentos semelhantes entre elas. Aliado a isso, técnicas de visualização da informação podem ser exploradas para auxiliar na detecção de padrões obtidos por estudantes nas diferentes áreas de conhecimento. Como forma de retornar *feedback* ao aluno, os tipos de perfis podem ser integrados ao software para uma classificação automática de apresentações em tempo real. Espera-se também testar outras técnicas de classificação para melhorar o desempenho do modelo.

Agradecimentos

O presente trabalho foi parcialmente realizado com apoio do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) por meio do Edital Universal 01/2016 processo 404369/2016-2 (Projeto: Comparação entre Diferentes Abordagens na Modelagem e Identificação de Acadêmicos em Risco em Cursos de Educação a Distância). Além do anterior, o trabalho também foi financiado pelo Projeto CORFO 14ENI2-26905 “Nueva Ingeniería para el 2030” - Pontificia Universidad Católica de Valparaíso, Chile.

Referências

- Chen, L., Leong, C. W., Feng, G., Lee, C. M., and Somasundaran, S. (2015). Utilizing multimodal cues to automatically evaluate public speaking performance. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*, pages 394–400. IEEE.
- Duda, R. O., Hart, P. E., and Stork, D. G. (1995). *Pattern classification and scene analysis* 2nd ed. ed: Wiley Interscience.
- Echeverría, V., Avendaño, A., Chiluíza, K., Vásquez, A., and Ochoa, X. (2014). Presentation skills estimation based on video and kinect data analysis. In *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge, MLA '14*, pages 53–60, New York, NY, USA. ACM.
- Freund, Y. and Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting.
- Junokas, M., Lindgren, R., Kang, J., and Morphey, J. (2018). Enhancing multimodal learning through personalized gesture recognition. *Journal of Computer Assisted Learning*.
- Le, T.-L., Nguyen, M.-Q., et al. (2013). Human posture recognition using human skeleton provided by kinect. In *Computing, Management and Telecommunications (ComMan-Tel), 2013 International Conference on*, pages 340–345. IEEE.
- Leong, C. W., Chen, L., Feng, G., Lee, C. M., and Mulholland, M. (2015). Utilizing depth sensors for analyzing multimodal presentations: Hardware, software and toolkits. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 547–556. ACM.
- Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137.
- Munoz, R., Villarroel, R., Barcelos, T., Souza, A., Merino, E., Guinez, R., and Silva, L. A. (2018). Development of a software that supports multimodal learning analytics: A case study on oral presentations. *Journal of Universal Computer Science*, 24(2).
- Ochoa, X. (2017). Multimodal Learning Analytics. In Lang, C., Siemens, G., Wise, A. F., and Gašević, D., editors, *The Handbook of Learning Analytics*, pages 129–141. Society for Learning Analytics Research (SoLAR), Alberta, Canada, 1 edition.
- Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., and Castells, J. (2018). The rap system: automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pages 360–364. ACM.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.
- Schneider, B. and Blikstein, P. (2015). Unraveling students' interaction around a tangible interface using multimodal learning analytics. *Journal of Educational Data Mining*, 7(3):89–116.
- York, D. (2013). *Investigating a Relationship between Nonverbal Communication and Student Learning*. PhD thesis, Lindenwood University.