

Using Multimodal Data to Find Patterns in Student Presentations

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

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.cechinel@ufsc.br, erick.merino@postgrado.uv.cl,

rodolfo.villarroel@pucv.cl, robson.lemos@ufsc.br, roberto.munoz@uv.cl

Resumo—Multimodal Learning Analytics is a subfield of Learning Analytics that uses data coming from complex learning environments and collected through alternative devices that are different from those normally observed in the Learning Analytics literature. The present work uses data captured by Microsoft Kinect and organized with Lelikëlen system to find patterns in students oral presentations during a given discipline. For that, a total of 16 different features related to the records of 43 students presentations (85 observations) were used to generate clusters of students with similar behavior. Initial results indicate three main different profiles of students according to their patterns in oral presentations: active, passive, and semi-active. Such findings can be further implemented in Lelikëlen system in order to allow instant feedback to students. Future work will also evaluate how students oral presentations patterns evolve during the semester, and compare patterns of students presentations across areas to evaluate whether there are similarities or not.

Index Terms—Multimodal learning analytics, clustering, data mining, students postures

I. INTRODUÇÃO

A linguagem corporal é uma peça importante nos processos de aprendizagem e comunicação. As posturas corporais e os gestos são utilizados em apresentações orais para transmitir ideias e mensagens. A comunicação entre pessoas por uma causa específica, como em uma apresentação, requer cuidado para transmitir o que se quer para a parte que está ouvindo ou olhando [1]. A captura, análise e utilização desse tipo de informação pertence a área de *Multimodal Learning Analytics* (MLA) [1].

Em sua essência o MLA visa alavancar dados de modalidades não tradicionais, de forma a estudar e analisar o aprendizado de estudantes em ambientes de aprendizado complexos [2]. Essas técnicas permitem que professores deem retorno aos estudantes nas relações de aprendizado. De outro lado, técnicas de inteligência artificial estão cada vez mais presentes, principalmente quando se tratam de grandes quantidades de dados.

A identificação e caracterização de estudantes é um ponto fundamental para melhorar os processos de ensino e aprendizagem. A avaliação dos estudantes pode variar em diferentes abordagens, mas os dados quase sempre são convertidos em dados computáveis. Uma vez que são convertidos, os dados

podem ser processados por uma coleção de algoritmos de Aprendizado de Máquina.

O MLA é uma área que cresce com a análise de dados de variadas fontes [3]. Os dados são capturados durante interações de aprendizado. As tecnologias atuais de multimídia e as técnicas de Aprendizado de Máquina progrediram a um ponto onde algoritmos de tempo real podem ser usados para processar vídeos, áudio, e outros materiais digitais. Isso produz características como posturas, gestos e *skeleton models* que são usados para dar apoio para o MLA. Bons comunicadores demonstram suas habilidades de apresentação com características verbais, e também através de características não verbais, como linguagem corporal, contato visual com a audiência, ou também pelo espaço que ele ocupa no palco [4].

O modelo generalizado pelo algoritmo de aprendizado de máquina usa dados passados para induzir uma hipótese. Assim, são necessários dados de estudantes em apresentações orais de forma a ser possível analisar essas apresentações. A captura desses dados pode ser por imagem, voz ou por outro meio. O termo Multimodal denota que diferentes formas de dados em conjunto são usados. Nesse sentido, as posturas de aprendizes em apresentações são de grande importância na análise de apresentações. As posturas de um aprendiz podem fornecer informações importantes sobre seu estado interior [5].

Existem certos tipos de problemas que têm resolução complexa por meio de programas, por exemplo, reconhecimento facial e de fala [6]. As técnicas de aprendizado de máquina, que criam modelos a partir de dados passados, são importantes em áreas onde existe a análise de grandes quantidades de dados. Aliar, o MLA com essas técnicas se mostra importante no sentido de auxiliar no desenvolvimento da educação e principalmente na análise do comportamento dos estudantes durante apresentações orais. Deste modo, um estudo para avaliar apresentações dos estudantes utilizando dados Multimodais foi realizado. As técnicas K-means e de Silhouettes foram utilizadas para agrupar e avaliar os resultados obtidos, respectivamente. Alguns padrões de comportamentos foram encontrados, estes estão relacionados com a atividade do estudante durante a apresentação oral. Diante do exposto, o objetivo principal do trabalho é encontrar padrões em apresentações orais de estudantes.

Este trabalho está organizado da seguinte forma: na Seção II, são apresentados os trabalhos que possuem alguma relação com este. Na Seção III, é apresentada a ferramenta utilizada na aquisição/tratamento dos dados, para o entendimento do trabalho. Já na Seção IV, é apresentada uma visão geral do sistema, além da descrição de todas as etapas realizadas para obtenção dos resultados. Os dados são descritos brevemente na Seção V. Na seção VI são apresentados os resultados e na seção VII são apresentadas as análises desses resultados. E por último, na Seção VIII, são apresentadas as considerações finais do trabalho a partir da realização do experimento, além dos trabalhos futuros pretendidos pelos autores.

II. TRABALHOS RELACIONADOS

Diferentes iniciativas foram propostas com o objetivo de gerar feedback para as pessoas. Por exemplo, no trabalho [3], foi usada uma abordagem de vídeo e Microsoft (MS) Kinect em conjunto para extrair os dados das apresentações. O objetivo do trabalho era avaliar a qualidade das apresentações orais. Para os dados de vídeo foram analisados contato visual. Desses dados, nove características de vídeo foram calculados, bem como a média do deslocamento horizontal da face. As características do MS Kinect, relacionadas com o movimento e postura dos estudantes foram também extraídas. Depois da etapa de pré-processamento dos dados e da seleção das características, o classificador *Weighted Logistic Regression* foi utilizado para avaliar os dados em separado.

Por outra parte, em [7], é apresentado uma abordagem utilizando o MS Kinect com o seu *Software Development Kit* (SDK). Foi usado essa ferramenta para resolver alguns problemas encontrados quando foram utilizadas apenas câmeras convencionais. Eles exploram a capacidade do uso de informação de *skeleton* que o Kinect oferece, deste modo pode-se reconhecer posturas humanas. O dispositivo pode reconhecer corretamente as posturas de deitar, sentar, levantar e se curvar. Para isso foi utilizado o classificador *Support Vector Machines* SVM.

Em um estudo sobre o desenvolvimento de um modelo automático de pontuação para apresentações em público usando interpretação multimodal [8], foram usados dados provenientes do Kinect para extrair as características do problema. Algoritmos de Aprendizado de Máquina, tais como SVM, *Random Forests* e *glmnet* foram usados para processar os dados. As características são baseadas em posturas, olhos, expressão facial, traços de movimento, entre outros.

Usando também um MS Kinect, o trabalho [9] faz uma comparação dos movimentos da mão e do pulso entre peritos e novatos enquanto estes completavam tarefas de engenharia. Eles encontraram que os peritos empregam muito mais ações com as duas mãos do que os novatos. Eles usaram k-means para agrupar os estudantes.

Usando dados do kinect para avaliar estratégias de estudantes interagindo com um TUI *Tangible User Interface*, o trabalho [10] determina, com algoritmos de agrupamento, que as posturas de estudantes ficam em três tipos de posições principais: ativo, a semi-ativo e o estado passivo. No trabalho

é descrito que o total do tempo gasto no estado ativo está relacionado com ganho de aprendizagem. De maneira oposta, o tempo gasto no estado passivo está relacionado com baixo ganho de aprendizagem.

Em [11] é realizada uma análise de identificação de indivíduos em dados antropométricos e de marcha. Os autores usam o Kinect para gerar os dados a serem analisados e classificadores, tais como KNN (*K-Nearest Neighborhood*) e SVM. Os resultados são considerados comparáveis a outros estudos do tema, com acurácia considerável. Utilizando IoT (*Internet of Things*) e MLA o trabalho [12] apresenta um estudo sobre o desenvolvimento da área de *Learning Analytics* com dados em ambientes de atividades físicas.

Alguns destes trabalhos estão relacionados com apresentações orais e outros se relacionam com outros tipos de ambientes complexos de aprendizagem. A relação destes trabalhos citados com o trabalho em questão se dá pelo uso de técnicas de aprendizado de máquina com o MLA. O presente trabalho apresenta uma abordagem para a identificação dos padrões utilizando as técnicas K-means e de silhouettes, técnicas que não foram utilizadas juntas nos trabalhos relacionados apresentados. Outra diferença é a base de dados utilizada, que foi extraída e pré-arranjada pelo software Lelikëlen [13].

III. FERRAMENTA LELIKËLEN

Para este trabalho foi utilizado o Lelikëlen [13]. Esta ferramenta permite detectar, armazenar e visualizar posturas corporais das pessoas gravadas. O software ainda possui as opções de agregar posturas personalizadas, exportar e importar cenas e um visualizador das posturas detectadas em função do tempo. A ferramenta, apresentada na Figura 1, ainda permite exportar os dados gerados para visualização com ferramentas de mineração de dados.

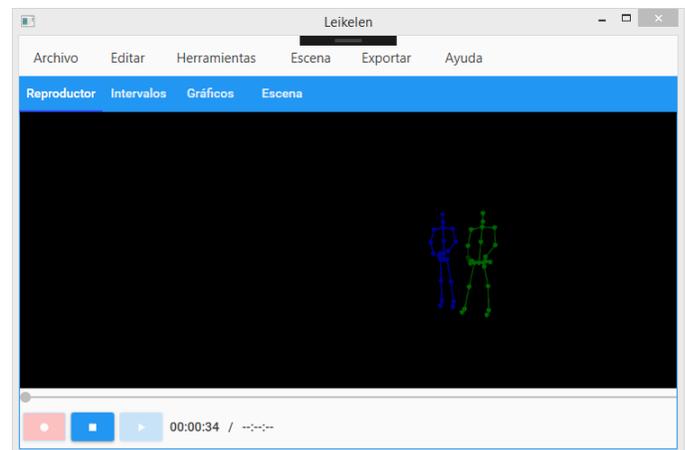


Figura 1: Tela do aplicativo Lelikëlen mostrando os modelos de *skeleton* de duas pessoas.

Para a captura de elementos corporais, a ferramenta utiliza o MS Kinect. Por um lado, para a classificação das posturas corporais utiliza o algoritmo ADABOOST [14], e tem 10 posturas

corporais previamente definidas e validadas [13]. Por outro lado, a ferramenta também possui outros tipos de métricas, como a distância proxêmica, que é calculada tomando como referência a distancia entre duas ou mais pessoas [15], ou se a pessoa está olhando para o público. A inclinação do corpo e se a pessoa está falando ou não.

IV. METODOLOGIA

Os procedimentos metodológicos são apresentados nesta seção. A forma como os dados foram apresentados, como os atributos se dividem e uma breve explicação que se refere aos dados utilizados é a etapa chamada de Dados. A segunda etapa é o Pré-processamento que consistiu na criação da base de dados utilizada. Com a criação da base de dados pode-se avaliar melhor as características, etapa chamada de Base de dados e características. Por fim, os algoritmos de Silhouettes e o K-means são aplicados à base de dados para posterior avaliação e compreensão dos resultados. A figura 2 apresenta a metodologia definida. Para a obtenção dos resultados, bem como todas as etapas de pré-processamento, foi utilizado o Matlab.

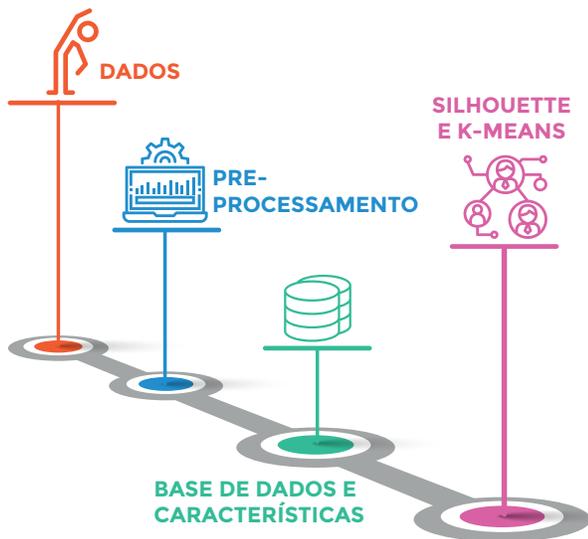


Figura 2: Esquema da metodologia.

A. Dados

A base de dados do problema em questão era composta por dois níveis de dados. O primeiro nível, com dados mais gerais, tem relação com características mais genéricas, como Postura Discreta e Ângulo da Cabeça. Estes dados ainda podem ser divididos em outro nível, chamado de dados específicos. Estes dados específicos tem origem nos dados do gerais, sendo uma subdivisão. A tabela I apresenta essas divisões. Como exemplo, o dado geral Postura Discreta foi dividido em outras 7 características específicas.

Um estudo de caso foi conduzido com estudantes universitários. Como parte da disciplina, foi solicitado aos alunos a realização de apresentações orais de projetos pré-definidos

dentro de três tópicos. Essas apresentações foram realizadas no mesmo ambiente onde os dados foram capturados [13].

Esses dados foram capturados com o Microsoft Kinect, além da classificação feita pelo Lelikeñen. O Kinect pode reconhecer mais de seis pessoas simultaneamente [16]. Foram realizadas 43 apresentações em duplas (exceto um aluno que decidiu apresentar sozinho), que foram sorteadas aleatoriamente. Ressalta-se que para este estudo, cada observação é um estudante em uma apresentação. Dessa forma, o objetivo principal do trabalho é encontrar padrões nessas características, sendo as apresentações o ambiente de aprendizado. Isso resultou em uma base de dados com 85 observações. Uma importante alteração feita nos dados apresentados na tabela I foi a exclusão da análise de 2 atributos que foram previamente coletados. Os atributos específicos Sentado e Pedindo Ajuda não aparecem nas análises dos resultados, bem como não foram usados como entrada para o algoritmo K-means. O motivo para essa exclusão foi que estas posturas não acontecem neste ambiente de aprendizado (apresentações orais), mas sim em aulas expositivas principalmente. Assim, decidiu-se que para este trabalho esses dados não seriam analisados.

B. Pré-processamento

Uma vez que os dados foram capturados e os atributos foram classificados pelo software Lelikeñen a etapa de Pré-processamento pôde ser realizada. Com os intervalos de cada ocorrência dos dados computados para cada um dos atributos avaliados (Mãos Baixas, Braços Cruzados entre outros). Dessa forma, foi realizada a soma total de cada intervalo para cada um dos atributos. Os intervalos em questão eram uma parte do total da apresentação. Dessa forma, o máximo possível para um determinado exemplo para um atributo é 1, que equivale a 100% do tempo da apresentação. Para cada exemplo da base de dados, foi computado a soma de seus dados (características) analisados em uma apresentação.

C. Base de dados e características

A base de dados ficou composta por 16 características (explicado na seção V). Estas foram extraídas dos dados das apresentações dos estudantes. Tipicamente, um exemplo, ou observação E , é representado por uma tupla de valores de atributos $(x_1; x_2; \dots; x_n)$, onde x_i é um valor de um atributo X_i . Para este problema, um exemplo é um estudante em uma apresentação. Já seus atributos são seus dados (explicados na seção IV-A) relacionados com comportamento em apresentações. Como não tem-se as classes para realizar a classificação dos exemplos, este é um problema chamado de aprendizado não supervisionado. O objetivo principal do estudo é explorar os dados e encontrar padrões em apresentações orais.

D. Silhouette e K-means

1) *Silhouette*: Para explorar os dados e analisar os grupos criados pelo K-means, o algoritmo de análise visual Silhouette [17] foi usado. Esse método é útil para selecionar o melhor número de clusters pré-definidos, por exemplo. Ele também

Tabela I: Descrição de cada atributo da base de dados inicial.

Geral	Específicas	Descrição
Voz	Fala	Detecta a voz da fonte
Postura discreta	Pedindo ajuda	Levanta uma mão para pedir ajuda
	Braços cruzados	Cruzou os dois braços
	Mãos na face	Tem uma mão no queixo
	Mãos na cabeça	Tem uma mão na nuca
	Mãos no quadril	Tem as mãos na cintura
	Mãos baixas	Tem as mãos para baixo
	Uma mão	Está explicando com uma mão (uma mão para baixo e a outra dobrada em posição explicativa)
	Mãos abertas	Está explicando com as duas mãos (as duas mãos com os braços dobrados)
	Apontando	Está apontando com uma mão (braço estendido horizontalmente)
Sentado	Está em posição sentado	
Ângulo da cabeça	Olhando o público	Está olhando o público
Proxemic	Público	A distância entre o ponto médio (externo) dos que apresentam é maior ou igual a 85cm
	Social	A distância entre o ponto médio (externo) dos que apresentam é maior ou igual a 65 e 85cm
	Pessoal	A distância entre o ponto médio (externo) dos que apresentam é maior ou igual a 45 e 65 cm
	Íntimo	A distância entre o ponto médio (externo) dos que apresentam é menor que 45cm
Proxemic Acurado	Público	Distância entre os pontos mais próximos daqueles que apresentam maior ou igual a 85cm
	Social	Distância entre os pontos mais próximos daqueles que apresentam está entre 65 e 85cm
	Pessoal	Distância entre os pontos mais próximos daqueles que apresentam está entre 45 e 65cm
	Íntimo	Distância entre os pontos mais próximos daqueles que apresentam é menor que 45cm
Inclinado	Ereto	A inclinação da pessoa está entre -0,333 e 0,333, com -1 inclinada para trás e 1 inclinada para a frente.
	Para cima	A inclinação da pessoa é menor que -0,333, com -1 inclinada para trás e 1 inclinada para a frente.
	Para baixo	A inclinação da pessoa é maior que 0,333, com -1 inclinada para trás e 1 inclinada para a frente.

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 [17]. 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.

2) *k-means*: O *k-means* [18] é um algoritmo iterativo de particionamento de dados que atribui i observações para exatamente um dos k clusters definidos pelos centroides. k é definido antes do algoritmo iniciar. Neste trabalho, o algoritmo *K-means* de agrupamento pode particionar os dados dos estudantes em k clusters. O agrupamento é alcançado minimizando a soma distâncias dos quadrados entre os dados e os centroides dos clusters.

V. DESCRIÇÃO E ANÁLISE DOS DADOS

A base de dados inicial utilizada na análise foi composta por 20 características. Foram consideradas as posturas mais específicas na análise. Isso se deve ao fato de que essas posturas fornecem mais informações para avaliação, como a posição das mãos, por exemplo. Após uma análise prévia dos dados observou-se que as posturas específicas, provenientes das posturas gerais Proxemic e Proxemic Acurado estavam com comportamento quase idêntico. Dessa forma, optou-se por retirar um desses conjuntos de posturas, nesse caso escolheu-se o Proxemic Acurado. Assim, como mencionado anteriormente, a base de dados utilizada foi composta de 16 atributos.

A figura 5 mostra o comportamento dos atributos Olhando o Público e de Fala. Nota-se que todos os exemplos da base foram inseridos em um gráfico, deste modo pode-se observar o comportamento geral. Esses atributos foram escolhidos devido as suas características distintas. Fica claro que esses dados são linearmente separáveis (podendo ser separados por uma reta no plano). Ressalta-se que a maioria dos atributos não possui essa características. Pela figura, pode-se visualizar claramente duas regiões de comportamento distinto. Parte das observações da base de dados estão mais para a parte superior do gráfico, já outra parte das observações estão posicionadas mais pela parte de baixo. Isso evidencia a separação entre essas duas variáveis.

A tabela II apresenta os valores totais de cada atributo usado no problema em todas as apresentações. Algumas características podem ser notadas com essa tabela. A presença de algumas variáveis é mais frequente que outras nas apresentações dos estudantes. Por exemplo, tem-se a característica Ereto que acontece com mais frequência que todas as outras. Outro ponto a se destacar é que a variável Íntimo ocorre poucas vezes ao longo das apresentações, assim como a postura Social. A variável Público também foi uma das posturas corporais bastante presente. Nota-se ainda um tempo alto geral olhando para o público e também com a postura Mãos Baixas.

Tabela II: Tempos totais de cada característica.

Atributo	Tempo Total[h:m:s]
Braços Cruzados	01:44:21
Para Baixo	00:00:58
Mãos na Face	00:27:20
Mãos a Cabeça	00:00:57
Mãos na Cintura	00:14:41
Mãos Baixas	03:21:20
Íntimo	00:00:38
Pessoal	00:00:01
Público	06:20:54
Social	00:00:14
Uma mão	00:49:27
Mãos Abertas	01:29:44
Apontando	00:45:02
Ereto	06:33:46
Fala	01:34:37
Olhando o Público	03:32:34

VI. RESULTADOS

A. Silhouette

Neste trabalho, foi utilizado o valor da silhouette para encontrar o melhor valor de k para o algoritmo K-means. Como cada observação possui um valor de silhouette $s(i)$, para encontrar o melhor valor, bastou-se realizar a média desse valores. Ressalta-se que mesmo que um determinado número de k apresente o melhor resultado, o mais importante é a avaliação posterior dos centroides. O melhor resultado em termos de valor médio foi o de $k=2$, apresentado na figura 3a, com os valores das silhouettes no eixo horizontal e os clusters no eixo vertical.

Como mencionado, outros números de clusters foram testados. A tabela III apresenta os valores médios conforme o valor de k muda. Um resultado interessante é o valor médio das silhouettes quando $k=3$. Este é o segundo melhor resultado do problema, significando que existem três diferentes grupos de comportamentos nas apresentações dos estudantes. A separação em dois grupos ($k=2$) não deixa margem para interpretação necessária, visto que as características de determinados grupos foram dissolvidas dentro de outros.

Assim, o melhor valor para a separação em clusters foi de $k=3$. Isso foi possível de visualizar após a realização da interpretação dos centroides. Ficou evidente que alguns comportamentos (análise realizada na seção 3b) das características não estavam aparecendo quando o número de clusters eram 2. Deste modo decidiu-se utilizar o valor de $k=3$ para a análise.

Tabela III: Comparação de diferentes valores de k para as Silhouettes.

Nº de Clusters	Média dos valores da Silhouette
2	0.415
3	0.333
4	0.280
5	0.206
6	0.191
7	0.203

Ainda nesta análise, a figura 3 mostra os valores das silhouettes para $k = 2, 3$ e 4 . É possível visualizar, quando

$k=4$, e por isso a técnica de silhouette é importante, que alguns exemplos nos clusters 1 e 4 possuem valores negativos. Isso significa que eles não pertencem aos seus respectivos clusters [17]. Ainda nota-se que para $k=4$, no cluster 4 existem poucos exemplos. Essas características indicam um número errado de clusters. Esse comportamento aparece quando o valor de k sobe (tabela III), para $k=5, 6$ e 7 .

B. K-means

Depois de procurar o melhor k para o algoritmo K-means, os clusters dos centroides foram avaliados. A avaliação dos centroides visa encontrar padrões nas características das observações. Desse modo, o comportamento dos clusters podem ser evidenciados. A figura 4 mostra os valores dos centroides de cada cluster, para cada um dos atributos avaliados.

Neste trabalho, após ser utilizado os algoritmos de Silhouettes e K-means, foram encontrados três padrões nos dados analisados. Os centroides, provenientes do K-means, foram analisados, e com estes foi possível identificar as características distintas de cada grupo. Pode-se visualizar pela figura 4 os grupos com os valores dos centroides no eixo horizontal.

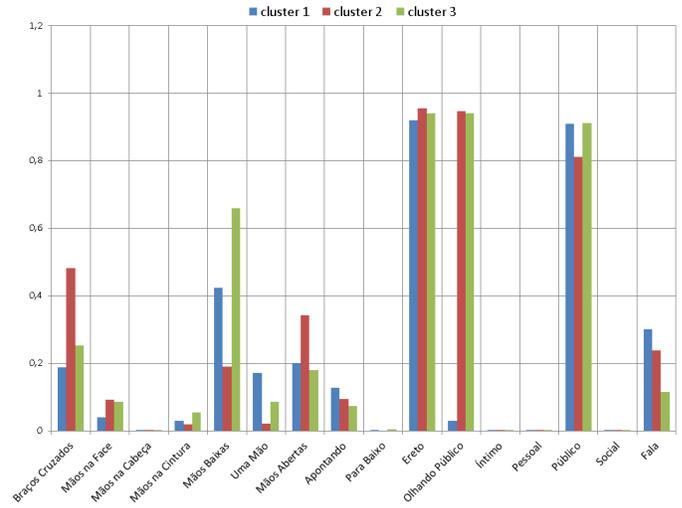


Figura 4: Valores dos centroides de cada característica avaliada.

É possível evidenciar, por exemplo, que o grupo 1 possui alguns atributos diferenciados em relação aos outros dois grupos. Primeiramente, pode-se visualizar que esse grupo é o que menos cruza os braços. Outra característica que destaca esse grupo é não olhar para o público. Nesta variável este grupo tem comportamento muito distinto dos demais grupos, que têm um valor alto, próximo de 1. O atributo de fala para o grupo 1 também é mais alto em relação a $k2$ e principalmente a $k3$. O grupo 1 também é maioria no conjunto de dados, são 41 exemplos de um total de 85 indivíduos.

O grupo 2 tem comportamento muito diferente de $k1$ e $k3$ na postura de braços cruzados. Este é o grupo que apresenta valor maior para esta variável, sendo praticamente o dobro do que os outros dois. Este grupo é o que menos fica com as mãos

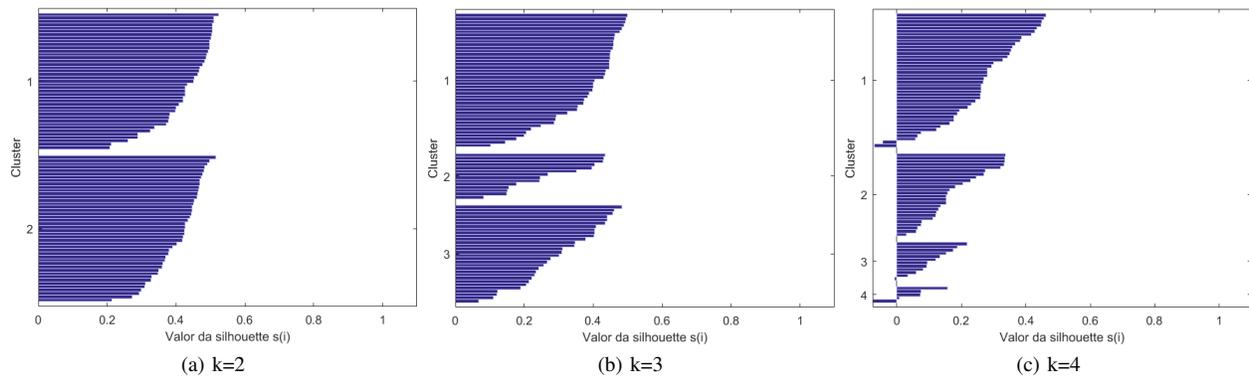


Figura 3: Gráficos das Silhouettes para $k=2, k=3$ e $k=4$.

baixas e também o que menos fica com uma mão, esta que tem relação com intenção de explicar algo para o público. Uma variável interessante deste grupo é a postura mãos abertas, que pode significar que a pessoa está explicando com ambas as mãos. Este grupo apresenta o maior valor para esta variável. Este grupo olha bastante para o público também, similar ao grupo 3. A característica de fala deste grupo é maior que a do grupo 3 e menor do que o 1. O grupo 2 é o menor grupo, apresentando 14 exemplos do conjunto de dados.

O terceiro grupo encontrado na análise tem características similares ao grupo 2. A característica de olhar para o público é muito similar entre esses grupos. A postura de mãos baixas é a maior entre todos os três, característica que está associada a um comportamento geralmente passivo. Este grupo também é o que menos aponta, mas fica mais com uma mão do que o grupo 2 e menos do que o grupo 1. Por fim, este grupo é o que menos fala entre todos, falando menos do que a metade do que o grupo 1. Este grupo é o segundo maior, possuindo 30 observações do conjunto total de dados.

Tabela IV: Quantidades e porcentagens de observações por rótulo.

	Nº de observações	Porcentagem [%]	Rótulo
Cluster 1	41	48,23%	Ativo
Cluster 2	14	16,47%	Semi-Ativo
Cluster 3	30	35,29	Passivo

Como o grupo 1 possui características geralmente associadas a um comportamento ativo, tais como a fala, a postura de apontar e ficar com uma mãos, este grupo posturas foi denominado de estado ativo. O grupo 3 possui características associadas a posturas com pouca ação, como as posturas de mãos baixas e ficar apontando, este conjunto foi denominado de estado passivo. Já o grupo 2 possui características marcantes do grupo um, como a fala e ficar apontando. Ainda, este grupo tem a postura de braços cruzados (postura relacionada à inatividade) muito alta e também fica com as duas mãos explicando. Pela similaridade com o grupo 1 e pelas suas posturas peculiares, decidiu-se chamar esse grupo de posturas de estado semi-ativo.

A tabela IV apresenta a quantidade de indivíduos em cada

cluster bem como o nome de cada grupo encontrado através dos padrões dos dados. Pode-se observar que o grupo 1 possui quase 50% do total dos dados. Estes rótulos foram colocados nesses conjuntos de atributos com base em [10], que usa uma abordagem semelhante, num contexto diferente, para encontrar padrões em ambientes de aprendizado complexo em estudantes.

A figura 5 mostra o resultado dos clusters nas variáveis de Fala e Olhando o Público. É possível visualizar os clusters para cada um dos exemplos do problema. Percebe-se que os padrões encontrados inicialmente, antes da aplicação do k-means, se confirmou. Os dados foram divididos pelos clusters nas regiões observadas. Ressalta-se o fato da presença do terceiro cluster apenas na região do cluster 2. Vale lembrar que em outros atributos, como o de fala, o cluster 3 tem comportamento mais parecido com o cluster 1.

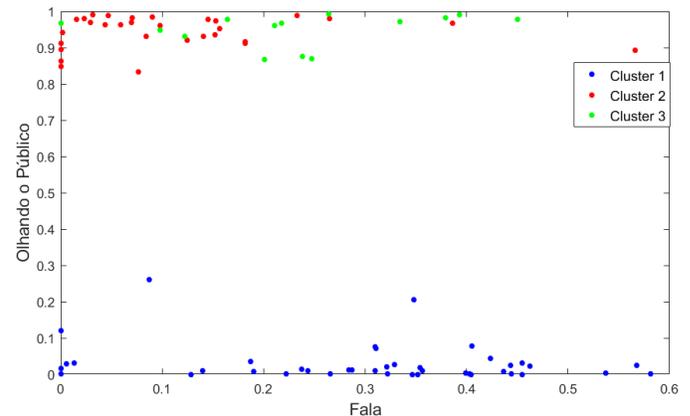


Figura 5: Posicionamento dos grupos de acordo com os atributos.

Os resultados apresentados no trabalho, como os grupos encontrados podem ser usados para retornar uma avaliação para os estudantes. Ainda mais especificamente, poderia analisar-se diretamente os valores dos centroides para retornar aos estudantes seu comportamento ao longo da apresentação. Da mesma maneira, os professores poderiam ter acesso a essas

ferramentas de análise, incorporando seu conhecimento com os resultados previamente encontrados.

O uso de outra base de dados, com outros exemplos e ainda com outros atributos avaliados poderia retornar um comportamento diferente dos dados. Um novo espaço de dados, com outros tipos de análises poderiam melhorar a distinção entre os rótulos.

VII. ANÁLISE E DISCUSSÃO

Neste trabalho foram encontrados três padrões nas apresentações dos estudantes: um padrão, chamado de Ativo, relacionado com uma ocorrência de mais fala do estudante e de posturas corporais mais participativas em apresentações como o atributo Apontando; Outro comportamento chamado de estado Passivo, mais relacionado com características de pouca participação nas apresentações, como os braços baixos; Ainda um terceiro padrão, chamado Semi Ativo, relacionado com o padrão ativo, mas com presença do comportamento passivo em suas características. A tabela V sumariza os resultados gerais da análise. Ainda foi possível mostrar a importância da análise dos centroides dos clusters. Essa análise possibilitou que fosse escolhido o valor de $k=3$, e não $k=2$, que tinha melhor resultado mas não mostrava todas os comportamentos e padrões presentes nos dados das apresentações.

Tabela V: Características gerais de cada cluster.

Rótulo	Características gerais
Ativo	Este grupo possui características geralmente associadas a comportamentos ativo. Estes exemplos são os que menos cruzam os braços e os que mais falam, o que mais apontam, indicando que está explicando algo e o que mais fica com uma mão, ainda são o grupo que não olham para o público.
Semi Ativo	Estes exemplos possuem tanto características do grupo ativo, quanto do grupo passivo. Este grupo olha muito o público, é o que mais cruza os braços (atributos Passivos). Fica muito com as mãos abertas, característica marcante. Fala menos que o grupo Ativo e mais que o grupo Passivo, por essa característica de fala decidiu-se que este seria semi Ativo.
Passivo	Similar ao grupo Semi Ativo estes exemplos também olham bastante para o público, é o que menos fala entre os três. É o que mais fica com as mãos baixas. Também é o que menos aponta entre os três. Fica com as mãos na cintura mais que os outros dois grupos. Essas características são geralmente relacionadas com um comportamento passivo.

VIII. CONSIDERAÇÕES FINAIS

Ambientes de aprendizado complexo têm sido cada vez mais estudados. O MLA auxilia nesse tipo de análise, uma vez que usa dados de diversas fontes como: vídeo, texto, áudio, gestos entre outros. Neste estudo, foram apresentados resultados de dados Multimodais acerca de apresentações orais de estudantes. As análises foram todas realizadas com a ajuda do Matlab.

Os padrões encontrados nos dados das apresentações foram o ativo, passivo e semi-ativo. Estes estão relacionados com os gestos e fala, além das posturas feitas pelos estudantes durante as apresentações. Dentre esses padrões, o ativo foi o

de maior ocorrência, com 41 exemplos pertencentes a ele. Os outros dois ficaram com 14 e 30, para o semi-ativo e passivo respectivamente.

Ressalta-se que a precisão das classificações da ferramenta Lelikëlen pode ter afetado o resultado da clusterização. Um número maior de observações pode melhorar o valor das silhouettes e ainda evidenciar os padrões encontrados neste trabalho. Mesmo assim, os resultados encontrados dão uma visão do comportamento geral dessas observações neste contexto.

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 essa análise entre áreas de conhecimento, uma análise temporal também pode revelar comportamentos distintos. As apresentações ao longo do semestre pode revelar a melhora do estudante em suas habilidades de apresentações orais, uma vez que vai dominando o conceito. Pode-se ainda avaliar num período maior, de um curso inteiro, e assim analisar os possíveis comportamentos encontrados. A técnica de *Sequential Pattern Mining* também será implementada visto que pode retornar resultados diferentes dos encontrados nesse trabalho.

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

- [1] C. W. Leong, L. Chen, G. Feng, C. M. Lee, and M. Mulholland, “Utilizing depth sensors for analyzing multimodal presentations: Hardware, software and toolkits,” in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 2015, pp. 547–556.
- [2] M. Worsley, “Multimodal learning analytics’ past, present, and, potential futures,” *International Conference on Learning Analytics & Knowledge*, 2018.
- [3] V. Echeverría, A. Avendaño, K. Chiluiza, A. Vásquez, and X. Ochoa, “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*, ser. MLA ’14. New York, NY, USA: ACM, 2014, pp. 53–60. [Online]. Available: <http://doi.acm.org/10.1145/2666633.2666641>
- [4] D. York, “Investigating a relationship between nonverbal communication and student learning,” Ph.D. dissertation, Lindenwood University, 2013.
- [5] X. Ochoa, “Multimodal Learning Analytics,” in *The Handbook of Learning Analytics*, 1st ed., C. Lang, G. Siemens, A. F. Wise, and D. Gašević, Eds. Alberta, Canada: Society for Learning Analytics Research (SoLAR), 2017, pp. 129–141. [Online]. Available: <http://solaresearch.org/hla-17/hla17-chapter1>

- [6] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification and scene analysis 2nd ed," ed: *Wiley Interscience*, 1995.
- [7] T.-L. Le, M.-Q. Nguyen *et al.*, "Human posture recognition using human skeleton provided by kinect," in *Computing, Management and Telecommunications (ComManTel), 2013 International Conference on*. IEEE, 2013, pp. 340–345.
- [8] L. Chen, C. W. Leong, G. Feng, C. M. Lee, and S. Somasundaran, "Utilizing multimodal cues to automatically evaluate public speaking performance," in *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*. IEEE, 2015, pp. 394–400.
- [9] M. Worsley and P. Blikstein, "Towards the development of multimodal action based assessment," in *Proceedings of the third international conference on learning analytics and knowledge*. ACM, 2013, pp. 94–101.
- [10] B. Schneider and P. Blikstein, "Unraveling students' interaction around a tangible interface using multimodal learning analytics," *Journal of Educational Data Mining*, vol. 7, no. 3, pp. 89–116, 2015.
- [11] V. O. Andersson and R. M. de Araújo, "Person identification using anthropometric and gait data from kinect sensor," in *AAAI*, 2015, pp. 425–431.
- [12] R. Martinez-Maldonado, V. Echeverria, O. C. Santos, A. D. P. D. Santos, and K. Yacef, "Physical learning analytics: a multimodal perspective," in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. ACM, 2018, pp. 375–379.
- [13] R. Munoz, R. Villarroel, T. S. Barcelos, A. Souza, E. Merino, R. Guiñez, and L. A. Silva, "Development of a software that supports multimodal learning analytics: A case study on oral presentations," *JUCS - Journal of Universal Computer Science*, vol. 24, no. 2, pp. 149–170, feb 2018.
- [14] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," 1995.
- [15] E. T. Hall, R. L. Birdwhistell, B. Bock, P. Bohannon, A. R. Diebold, M. Durbin, M. S. Edmonson, J. L. Fischer, D. Hymes, S. T. Kimball, W. L. Barre, S. J. Frank Lynch, J. E. McClellan, D. S. Marshall, G. B. Milner, H. B. Sarles, G. L. Trager, and A. P. Vayda, "Proxemics [and comments and replies]," *Current Anthropology*, vol. 9, no. 2/3, pp. 83–108, 1968.
- [16] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE multimedia*, vol. 19, no. 2, pp. 4–10, 2012.
- [17] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.
- [18] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.