

## APLICAÇÃO DE APRENDIZAGEM DE MÁQUINA EM DADOS CAPTURADOS PELO KINECT™: O ESTUDO DE CASO DO RECONHECIMENTO DE SINAIS DE LIBRAS

**SIMONE RUTZ<sup>1</sup>; LUCAS TORTELLI<sup>2</sup>; MARILTON AGUIAR<sup>3</sup>**

<sup>1</sup>Universidade Federal de Pelotas – sdrutz@inf.ufpel.edu.br

<sup>2</sup> Universidade Federal de Pelotas – lmtortelli@inf.ufpel.edu.br

<sup>3</sup> Universidade Federal de Pelotas – marilton@inf.ufpel.edu.br

### 1. INTRODUÇÃO

Segundo o último censo realizado no ano de 2010, cerca de 5,1% da população brasileira são surdos ou possuem algum grau de deficiência auditiva IBGE (2010). Dessa forma a necessidade por inclusão é crescente, principalmente em se tratando da interpretação da linguagem usada por eles que se difere em vários aspectos da linguagem natural. Portanto, abordagens para o reconhecimento de gestos de LIBRAS (Língua Brasileira de Sinais) fazendo uso de tecnologias assistivas são necessárias. Até então, os custos, tanto de equipamento quanto de processamento, tornam essas tecnologias pouco viáveis e acessíveis no dia a dia.

Mais recentemente, têm surgido novas tecnologias para reconhecimento de movimentos, como por exemplo o Microsoft Kinect™, utilizado inicialmente para a interação com jogos. Tal tecnologia se mostrou útil também para estudos que envolvem o reconhecimento de LIBRAS, pois além de coletar uma vasta gama de dados, como por exemplo pontos que mapeiam completamente o esqueleto humano, ele também é uma forma de tecnologia acessível e que não exige muito do usuário, pois este só precisa se posicionar em frente ao Kinect™.

A integração de técnicas de mapeamento, reconhecimento de padrões e aprendizagem de máquina pode ser utilizada para interpretação de gestos específicos. É neste contexto que o projeto LIKI – *Libras e Kinect Aplicados às Tecnologias Assistivas*, em desenvolvimento pelo grupo de pesquisa, está inserido. O projeto busca propor modelos e ferramentas computacionais para auxiliar no processo de ensino-aprendizagem da Língua Brasileira de Sinais (LIBRAS) para crianças com algum tipo de deficiência auditiva.

Mais especificamente, uma das temáticas abordadas pelo projeto trata do desenvolvimento e a aplicação de técnicas de reconhecimento de padrões para a identificação dos gestos. Outras temáticas abordadas no projeto envolvem capturar, especificar e armazenar de modo apropriado os movimentos do usuário utilizando o sensor Kinect™ (versão 2) e, então, integrá-las em um software que será utilizado em sala de aula.

Assim, este artigo pretende apresentar os resultados preliminares da busca do melhor algoritmo de treino e reconhecimento de sinais de LIBRAS, aqueles mais apropriados de acordo com os dados coletados fazendo uso do Kinect™. Para isso, optou-se pelo uso da plataforma Weka (2013) amplamente utilizada na área de reconhecimento de padrões, desenvolvida pela Universidade de Waikato, que é uma coletânea algoritmos de aprendizado de máquina (SMOLA, 2008). Esta ferramenta conta com vários métodos de aprendizagem e, além disso, possui o seu próprio formato de arquivo que é o ARFF, uma variante do formato CSV. Dentre a vasta gama de métodos disponibilizados pelo WEKA, para estes trabalhos, foram selecionados para avaliação os seguintes métodos, apresentados em WITTEN (2011): IBK (K-nn), J-48, Random Forest e, por fim,

Naive Bayes. Para fins de comparação, foram escolhidos os mesmos métodos utilizados em SOUSA (2012) (IBK e J-48), em ZHANG (2012) (*Random Forest*) e em GOMES (2012) (Naive Bayes).

Serão apresentados nas próximas Seções: a coleta dos dados e o tratamento dos mesmos; considerações sobre o Kinect™ justificando de forma mais adequada a sua escolha; e, uma breve descrição de cada algoritmo escolhido. Além disso, serão apresentados os resultados encontrados e as conclusões que foram obtidas até o momento.

## 2. METODOLOGIA

Para a construção da base de dados que foi usada nos testes foram escolhidos cinco gestos, que representam os dígitos de 0 a 5, na linguagem de LIBRAS. Cada um dos gestos foi repetido cinco vezes, por cinco pessoas diferentes, totalizando 125 gestos. Teve-se o objetivo de montar uma base consistente, capaz de prever as possíveis variações no momento de realizar qualquer uns desses gestos.

Sabe-se que o Kinect™ atualmente é capaz de mapear o corpo humano em diversos pontos de seu esqueleto, da cabeça aos pés SOUSA (2012). De seus pontos é possível pegar as coordenadas tridimensionais. As informações dos pontos foram coletadas e organizadas em um arquivo de dados do tipo ARFF, onde coordenadas e algumas informações de distâncias percorridas pelos pontos em movimento foram consideradas como atributos de interesse para a realização dos testes.

A segunda versão do Kinect™, utilizada neste projeto, possui uma câmera tridimensional (3D) possibilitando um maior alcance de profundidade e de cor; e, sua imagem tem três vezes mais precisão que a versão anterior do mesmo dispositivo. Essa câmera possui resolução de 1080p e a captura independe da luminosidade, uma vez que possui sensor infravermelho. Além de reconhecer 25 pontos do corpo, o Kinect™ ainda percebe três estados diferentes da mão: aberta, fechada e lasso, que constitui da mão parcialmente aberta, com apenas dois dedos visíveis (MICROSOFT, 2015).

Como mencionado anteriormente, os algoritmos escolhidos para os testes são amplamente utilizados no reconhecimento de gestos. O algoritmo IBK é uma implementação do algoritmo K-nn (*K-nearest-neighbor*) que consiste em encontrar os K-vizinhos mais próximos num conjunto de características. O J-48 também é uma variação de outro algoritmo conhecido, o C4.5, que consiste na geração de uma árvore de decisão, muito utilizadas em classificação.

Já o algoritmo *Random Forest* baseia-se na criação de uma floresta de árvores de decisão, bem semelhante ao J-48. E, por fim, o algoritmo Naive Bayes, é um classificador probabilístico que se baseia na aplicação do teorema de Bayes, pressupondo independência entre atributos. Ressalta-se que estes algoritmos foram os que melhor obtiveram resultados nos trabalhos de ZHANG (2012), SOUSA (2012) e GOMES (2012).

O Weka permite algumas configurações de parâmetros para a realização dos testes com relação à manipulação da base de dados, tais como: *Use training Set*, *Suplied test set*, *Cross-Validadion* e *Percentage Split* (WITTEN, 2005). Ao selecionar *Use training Set*, as fases de treino e de teste do método são realizadas com exatamente os mesmos dados da base. *Suplied test set* é uma opção que permite que após o treino o algoritmo seja testado com uma outra base de dados diferentes. O parâmetro *Cross Validation* permite criar subconjuntos dos dados e, então, faz a análise do melhor método a partir da média de seus

resultados. E, por fim, *Percentage Split*, onde parte da base é separada em treino e teste a uma porcentagem escolhida pelo usuário.

### 3. RESULTADOS E DISCUSSÃO

A base final de dados possui aproximadamente 100 instâncias e para a realização dos testes as configurações consideradas no Weka tem os valores padrões da própria ferramenta, a saber: *Cross Validation* em 10 subconjuntos e *Percentage Split* em 66%. Entre os parâmetros *Use training set* e *Supplied test set* observou-se que como não se dispunha de outra base de dados para aplicar o teste, ambos produziam os mesmos resultados. Logo, manteve-se os resultados obtidos usando a opção *Use training Set*.

O Weka além de aplicar os algoritmos a uma base de dados, ele retorna uma análise bem aprofundada dos resultados obtidos, com várias estatísticas associadas aos mesmos. Na Tabela 1, pode-se observar a porcentagem de acerto de cada algoritmo para cada configuração do Weka. Os resultados apresentados incluem a classificação dos dados usando três das quatro configurações possíveis do Weka.

Tabela 1 – Resultados da classificação dos dados.

Configuração	Algoritmos			
	IBK(Knn)	J48	Random Forest	Naive Bayes
Percentage Split (66%)	44,74%	21,05%	50%	23,68%
Cross-Validation (10)	45,95%	36,04%	<b>56,76%</b>	22,52%
Use training Set	100%	93,69%	100%	46,85%

Observa-se que o algoritmo que obteve melhores resultados no geral foi o *Random Forest*, mesmo assim bem abaixo do esperado. Em seguida, o método que melhor se adaptou foi o IBK, com valores consideravelmente próximos do encontrados pelo *Random Forest*. O J-48 seguido do Naive Bayes foram os métodos com pior desempenho.

Vale ressaltar que os resultados encontrados para a configuração *Use training set* não podem ser considerados como os melhores, pois testar o modelo com os mesmos dados que ele foi treinado não realmente prova que ele “aprendeu” de fato os gestos em questão. Isso deveu-se a pouca quantidade de gestos utilizados para gerar o banco de dados.

Para tentar melhorar os primeiros resultados, uma abordagem um pouco diferente foi considerada. Nesta abordagem, os testes foram repetidos com os mesmos parâmetros, mas aplicou-se o filtro *Discretize* na preparação dos dados.

O filtro *Discretize* realiza a discretização de uma faixa de atributos numéricos em um *dataset* de atributos nominais, como o objetivo de tornar os dados mais homogêneos. Entretanto, esta abordagem não obteve sucesso, produzindo resultados inferiores, como é apresentado na Tabela 2.

Tabela 2 – Resultados com o filtro *Discretize*.

Configuração	Algoritmos			
	IBK(Knn)	J48	Random Forest	Naive Bayes
Percentage Split (66%)	31,58%	34,21%	34,21%	21,05%
Cross-Validation (10)	37,84%	38,74%	37,84%	33,33%
Use training Set	63,96%	63,96%	63,96%	63,96%

E como pode ser observado, o desempenho dos algoritmos caiu. Mesmo assim, o algoritmo que mais se destacou continua sendo o *Random Forest* seguido pelo J-48 que desta vez supera o IBK e, repetindo o pior desempenho, o Naive Bayes.

Sobre estes resultados alguns pontos podem ser levantados e trazidos à discussão, principalmente em se tratando da base de dados. Levantou-se hipóteses de que possa ser necessário padronizar a posição do usuário em frente ao Kinect, através de uma calibração. Além disso, é válido considerar que nem todos os pontos coletados (neste trabalho foram utilizados os dados do esqueleto completo) sejam realmente necessários para o processo de reconhecimento dos gestos; e, por último e não menos importante, pode-se reavaliar os atributos que compõe a base.

#### 4. CONCLUSÕES

Trabalhar com o desenvolvimento de tecnologia assistivas é sempre um desafio, pois não envolve somente a aplicação de uma determinada tecnologia. Além disso, há sempre muito a ser considerado, a experiência do usuário e o modo como a tecnologia vai reagir a diferentes ambientes e pessoas que a utilizam.

A realização desses testes, mesmo com resultados não tão satisfatórios, foi fundamental para os próximos passos do projeto, sabe-se agora que não basta simplesmente coletar e utilizar todos os dados fornecidos pelo Kinect™, sem que haja um tratamento prévio dos mesmos. Pode-se também ser necessário restringir a quantidade de pontos do esqueleto a serem considerados para refinar a base de dados, ou seja, há muito a se contribuir nesse ponto.

Para o futuro, depois de se debruçar nos gestos estáticos, há o intuito de se trabalhar com gestos dinâmicos, pois o que se sabe é que LIBRAS tem sua própria sintaxe e semântica atribuídos. Assim, trabalhar com gestos dinâmicos tende a ser uma grande contribuição tanto para o mundo acadêmico, quanto principalmente para os usuários da Língua Brasileira de Sinais.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

- GOMES, P., MORGENS, S., SMITH, S. Gesture Classification from Kinect Data **CMPS242: MACHINE LEARNING**, Santa Cruz (CA), USA, 2012.
- IBGE. **Resultados do Censo de 2010**. Acessado em 20 jul. 2015. Online. Disponível em: <http://censo2010.ibge.gov.br/resultados>.
- MICROSOFT. Kinect for Windows features. Acessado em 22 jul. 2015. Online. Disponível em: <https://www.microsoft.com/en-us/kinectforwindows/meetkinect/features.aspx>
- SMOLA, A., VISHWANATHAN, S. V. N. **Introduction to Machine Learning**. Cambridge: Cambridge University Press, 2008, 1v.
- SOUSA, A. P. A. **Interpretação da Linguagem Gestual Portuguesa**. 2012. Dissertação (Mestrado em Engenharia Informática). Universidade de Lisboa.
- WEKA. **Data Mining Software in Java**. Acessado em 15 jul. 2015. Online. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>.
- WITTEN, I. H., FRANK, E. **Data Mining Practical Machine Learging Tools and Techniques**. San Francisco: Elsevier, 2005. 2v.
- WITTEN, I. H., FRANK, E., HALL, M. A. **Data Mining – Practical Machine Learging Tools and Techniques**. Burlington: Elsevier, 2011. 3v.
- ZHANG, Hao; DU, WenXiao; LI, Haoran. **Kinect gesture recognition for interactive system**. 2012.