Classificação de Expressões Faciais Negativas na Língua Brasileira de Sinais

Ana Gabriela P. e Silva[1], **Emanuel O. da Silva**[1], **Giovana de Lucca**[1], **Letícia C. Passos**[1], **Matheus M. Matos**[1], **Elloá B. Guedes** [1]

[1]Núcleo de Computação
Escola Superior de Tecnologia
Universidade do Estado do Amazonas (UEA)
Manaus – AM – Brasil

ebgcosta@uea.edu.br, {agps, eos, gol, lcps, mmmt}.eng@uea.edu.br

Abstract. *The Brazilian Sign Language is a way for the social opening of deaf, deafblind and other kinds of disabilities. This paper proposes a computational model capable to recognize negative facial expressions in Brazilian Sign Language. For this purpose, were used the Multiayer Perceptron Atificial Neural Networks developed on Python. Compared with other literature results, which used the same dataset for the same purpose, the developed Neural Networks obtained better results with an F-score about 77%.*

Resumo. *A Língua Brasileira de Sinais é um caminho para a abertura social de pessoas surdas, surdo-cegas e com outros tipos de deficiência. Este artigo propõe um modelo computacional capaz de reconhecer expressões faciais negativas na Língua Brasileira de Sinais. Para isso, foram utilizadas Redes Neurais Atificiais do tipo Multilayer Perceptron deselvolvidas na plataforma Python. Em comparação com outros resultados da literatura, que utilizou o mesmo conjunto de dados para o mesmo objetivo, as Redes Neurais desenvolvidas obtiveram resultados melhores com um F-score em torno de 77%.*

1. Introdução

A *Língua Brasileira de Sinais* (Libras) é uma linguagem gesto-visual segundo a qual emite-se uma mensagem por meio de movimentos nas mãos, corpo e face e quem recebe a mesma o faz com os olhos [de Fátima Brecailo 2012]. Esta língua é um caminho para a abertura social de pessoas surdas, surdo-cegas e com outros tipos de deficiência, tendo sido reconhecida pela Lei Federal 10.436 de 24/04/2002 [Flores et al. 2012].

Em Libras, cada palavra é representada por um sinal e carrega níveis linguísticos diferentes, tais como fonologia, morfologia, sintaxe e semântica. Além disso, as expressões faciais e corporais corroboram para o entendimento da informação. As expressões faciais, em particular, pode ser afetivas ou gramaticais. Quando possuem caráter gramatical, as chamadas *expressões faciais gramaticais*, conferem informação gramatical a uma sentença expressa em sinais, complementando o seu sentido. Estas expressões podem ser de nove tipos gerais: afirmativa, negativa, condicionais, relativas, com tópico, com foco, interrogativa binária (questões do tipo 'sim' ou 'não'), interrogativa de dúvida e interrogativa qu (quando, que, onde, como, etc.) [de Almeida Freitas et al. 2014b].

Para automatizar o reconhecimento de Libras, é essencial, portanto, a análise das expressões faciais gramaticais. Em consequência, alguns outros trabalhos da literatura já exploraram esta perspectiva. O trabalho de Freitas et al. [de Almeida Freitas et al. 2014b] considerou o reconhecimento das oito expressões faciais gramaticais por meio da utilização de métodos da Aprendizagem de Máquina (AM). Para tanto, os autores conceberam um conjunto de dados, intitulado *Grammatical Facial Expressions Data Set*, que contém 100 coordenadas de diferentes posições faciais obtidas a partir de um sensor e anotadas de maneira supervisionada enquanto um sujeito emitia sentenças em Libras [de Almeida Freitas et al. 2014a].

Utilizando apenas 17 coordenadas (x, y) dos 100 pontos disponíveis na face, os autores treinaram e testaram uma rede neural artificial para o problema de classificação da expressão facial gramatical correspondente. A rede considerada, do tipo *multilayer perceptron* com 10 neurônios na camada oculta, foi treinada com diferentes taxas de aprendizado. De acordo com os resultados obtidos, percebeu-se que algumas expressões foram classificadas corretamente em 91% das instâncias dos casos de testes. Porém, as expressões negativas não foram bem classificadas, com *F-score* em torno de 45%.

Considerando as limitações identificadas para detectar expressões faciais gramaticais negativas, este trabalho se propõe a explorar outras redes neurais que possam melhor se adequar ao problema, objetivando o aumento nos índices de correta classificação deste tipo de expressão. Para apresentar os resultados obtidos, este trabalho está organizado como segue. Os conceitos fundamentais para o entendimento deste trabalho são apresentados na Seção 2. Uma visão geral do conjunto de dados encontra-se detalhada na Seção 3. A metodologia utilizada para conceber as diferentes redes neurais para este problema pode ser vista na Seção 4. Os resultados e a discussão são apresentados na Seção 5. Por fim, as considerações finais e sugestões de trabalhos futuros são mostrados na Seção 6.

2. Fundamentação Teórica

Nesta seção serão apresentados os fundamentos teóricos que dão suporte à realização deste trabalho. Esta seção inclui os conceitos elementares sobre Expressões de Gramática Facial e Redes Neurais *Multilayer Perceptron*.

2.1. Expressões de Gramática Facial

As expressões de gramática facial estão relacionadas a estruturas do nível da morfologia e do nível da sintaxe e são obrigatórias nas línguas de sinais em contextos determinados [Quadros and Karnopp 2009]. No nível morfológico, as expressões faciais correspondem ao grau de intensidade de um adjetivo ou ao grau de tamanho para um substantivo. Já no nível da sintaxe, as expressões determinam o tipo da estrutura da sentença, como pode ser visto na Tabela 1 a seguir.

Para realizar estas expressões devem ser utilizados o movimento da cabeça, a direção do olhar, a elevação das sobrancelhas, o franzir da testa, como também os movimentos dos lábios para indicar negação, como pode ser visto na Figura 2.1. Cada expressão possui características bem definidas e pode aparecer mais de uma vez em uma única sentença, entretanto a falta delas pode deixar uma frase sem sentido [Quadros and Karnopp 2009].

Tipo de sentenças	Exemplo
Afirmativa	Eu irei à escola
Negativa	Não gosto de chocolate
Condicional	Se chegarmos no horário iremos ao cinema
Interrogativa com pronome interrogativo	Quando é o seu aniversário?
Interrogativa binária (sim/não)	Você vai ao shopping hoje?
Interrogativa de dúvida	Você tem certeza que esse lápis é seu?
Relativa	O carro que quebrou está na oficina
Tópico	Cores, eu gosto de vermelho
Foco	O almoço foi risoto. Não, o almoço foi arroz

Tabela 1: Tipos de construções de sentenças e exemplos correspondentes.

(a) Negativa (b) Afirmativa (c) Interrogativa

Figura 1: Exemplos de Expressões de Gramática Facial. Fonte das imagens: [Sousa 2011]

Neste trabalho será abordado somente a expressão facial do nível de sintaxe, especificamente do tipo de construção negativa. A expressão de gramática facial negativa, segundo [Arroteia 2005], pode ser indicada de duas formas. Uma forma é realizado com o movimento para os lados, porém este movimento não é obrigatório em Libras e refere-se às questões discursivas. A outra forma é através da modificação do contorno da boca, juntamente com o abaixamento das sobrancelhas e levemente da cabeça. Esta última forma, porém, é obrigatória para marcar a negação, pois está relacionada às questões sintáticas.

2.2. Redes Multilayer Perceptron

Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados no cérebro humano e que possuem a capacidade de aquisição e manutenção de informações [Richter et al. 2007]. Uma das principais características das RNAs é aprender conforme o ambiente onde estão inseridas e assim melhorar seu desempenho [Souza and Monteiro 2009]. Além disso, as RNAs possuem a capacidade de generalização que permite a classificação de padrões de entrada anteriormente desconhecidos [Senger et al. 2007].

As RNAs são compostas de unidades de processamento, conhecidas como neurônios artificiais, dispostas em uma ou mais camadas [Faceli et al. 2011]. Os neurônios são elementos processadores interligados, trabalhando em paralelo para desempenhar uma determinada tarefa [Vellasco 2007]. Uma RNA recebe atributos de entrada que são ponderados e combinados por meio dos neurônios por uma função matemática

chamada função de ativação. A saída dessa função é a resposta do neurônio para a entrada [Faceli et al. 2011].

Para resolver problemas não linearmente separáveis utilizando RNAs, a alternativa mais utilizada é adicionar uma ou mais camadas ocultas [Faceli et al. 2011]. A RNA *Multilayer Perceptron* (MLP) é uma extensão do Perceptron simples, capaz de trabalhar com problemas não linearmente separáveis [Lima et al. 2011]. As redes do tipo MLP apresentam uma camada de entrada, uma ou mais camadas ocultas de neurônios e uma camada de saída [Faceli et al. 2011]. A função das camadas ocultas é extrair características apresentadas nos padrões de entrada permitindo que a rede crie sua representação, mais rica e complexa [Lima et al. 2011].

O algoritmo mais tradicional utilizado no treinamento de uma rede MLP é o algoritmo de retropropagação do erro ou *backpropagation*, conhecido também como regra delta generalizada [Soares and Teive 2015]. Esse algoritmo é baseado em gradiente descendente e necessita que a função de ativação seja contínua, diferenciável e, de preferência, não decrescente. Uma das funções que obedecem a esses requisitos é a função de ativação do tipo sigmoidal [Faceli et al. 2011]. A técnica *backpropagation*, porém, apresenta convergência muito lenta. Assim, outros métodos surgiram para minimizar essa deficiência, como, por exemplo, a utilização da taxa de aprendizado adaptativa [Wagner et al. 2013].

3. Visão Geral do Conjunto de Dados

O *dataset* escolhido para o trabalho foi o *Grammatical Facial Expressions Data Set*, obtido no repositório da UCI (*University of California, Irvine*) [Fernando de Almeida 2014]. Ele armazena os tipos de expressão facial gramatical: afirmativa, negativa, condicionais, com tópico, com foco, interrogativa binária (questões do tipo 'sim' ou 'não'), interrogativa de dúvida e interrogativa qu (quando, que, onde, como, etc.) de duas pessoas. Para cada expressão tem-se um *dataset* com os atributos de entrada e um com o atributo de saída. Os de entrada são ao todo 301, sendo eles um *timestamp*, que é o tempo do *frame* de vídeo em que a expressão foi feita, e as coordenadas x e y em pixels e z em mm de 100 pontos dispostos sobre a face. Os dados de saída contém um atributo que informa se os dados de entrada correspondem ao tipo de expressão do conjunto de dados, por exemplo considerando o conjunto da expressão negativa, se o atributo de saída for 1, isso significa que os dados de entrada correspondem a uma expressão negativa, caso contrário o atributo será 0, isso quer dizer que a expressão não é negativa, mas poderá ser afirmativa, interrogativa ou outras, ou até mesmo nenhuma dos tipos considerados.

Como foi proposto a melhoria do desempenho da classificação da expressão negativa, o único *dataset* considerado foi o que apresenta os dados referentes a ela. Os pontos considerados foram os mesmos do trabalho de [de Almeida Freitas et al. 2014b], sendo ao todo 17, os quais se localizam nas regiões: *left eye, right eye, left eyebrow, right eyebrow, nose, mouth, nose tip, line above left eyebrow, line above right eyebrow*.

Para fins de simplificação, os atributos de entrada e saída foram unidos em um único *dataset*, os dados da pessoa a e da pessoa b foram agrupados em um único conjunto e o *timestamp* foi desconsiderado, pois não influencia na obtenção do resultado. Além disso, os atributos correspondentes às coordenadas z também foram desconsiderados, pois além de possuírem erros de medição do sensor, uma futura aplicação por captura

de imagens poderia não conter coordenadas de profundidade. Por fim, os dados foram normalizados, exceto os do atributo alvo.

Uma análise mais detalhada dos dados se fez necessária e, portanto, os dados foram agrupados de forma que cada grupo correspondesse à uma região da face. A partir dos grupos formados foi realizada uma estatística descritiva, sintetizada na tabela 2.

AB-negativo	Pontos	media	mediana	desvio padrão
	left eye – x	303.44	304.93	11.33
	left eye – y	218.93	229.06	15.13
	right eye- x	332.18	333.15	11.23
	right eye – y	218.49	227.43	13.83
	left eyebrow – x	300.66	301.33	12.53
	left eyebrow – y	211.23	221.23	15.11
	right eyebrow – x	334.37	333.63	11.62
	right eyebrow – y	210.90	218.82	13.59
	nose – x	317.62	317.82	12.76
	nose – y	228.28	232.96	17.57
	mouth – x	318.39	319.05	14.18
	mouth – y	245.03	255.32	18.67
	face contour – x	319.83	319.82	27.16
	face contou – y	245.97	243.08	22.02
	left iris – x	303.44	306.49	10.69
	left iris – y	218.93	229.05	15.07
	right iris – x	332.18	334.14	10.56
	right iris – y	218.49	228.07	13.76
	nose tip – x	317.64	319.62	11.49
	nose tip – y	231.87	241.99	17.23
	line above left eyebrow – x	229.75	300.41	13.46
	line above left eyebrow – y	207.75	217.39	14.79
	line above right eyebrow – x	335.13	335.04	12.08
	line above right eyebrow – y	207.07	214.91	13.17

Tabela 2: Métricas de verificação

Analisando-se a tabela, é possível observar que as variáveis consideradas não estão balanceadas e possuem um alto desvio padrão. Isso ocorre em decorrência da alta dispersão verificada nos dados correspondentes ao indivíduo b.

4. Materiais e Métodos

Nesta seção serão apresentados os materias e métodos que foram utilizados para o desenvolvimento deste trabalho, bem como a métrica de desempenho empregada. Esta seção inclui os conceitos elementares sobre *F-score*, qual a arquitetura das RNAs utilizadas e a plataforma utilizada para o desenvolvimento dessas redes.

4.1. Métricas para avaliação de desempenho

Na análise estatística de classificação binária, o *F-Score* (F-medida) é uma medida de precisão de teste. A pontuação *F-score* pode ser interpretada como uma média ponderada

entre *precision* e *recall*, onde a pontuação atinge seu melhor valor em 1 e pior em 0.

Considerando a *precision* (precisão) p e o *recall* (recuperação) r para calcular a pontuação do teste: p é o número de exemplos classificados como pertencentes a uma classe, que realmente são daquela classe (verdadeiro positivo), dividido pela soma entre este número, e o número de exemplos classificados nesta classe, mas que pertencem a outras (falso positivo); r é o número de exemplos classificados como pertencentes a uma classe, que realmente são daquela classe, dividido pela soma entre este número e o número de exemplos que pertencem a outra classe, que realmente pertencem aquela classe. No caso binário, verdadeiros positivos divididos pela soma deste número e os número de falsos negativos.

$$\text{F-Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{1}$$

4.2. Arquiteturas das Redes

Para a escolha das arquiteturas considerou-se a função de ativação, taxa de aprendizado e *solver* padrão fornecidos pela biblioteca *scikit-learn*. Foram utilizadas uma e duas camadas ocultas e para verificação da quantidade de neurônios nas camadas ocultas foi utilizada a regra da pirâmide geométrica, descrita na equação 2, sendo N_h o número de neurônios na camada oculta, N_i o número de neurônios na camada de entrada e N_o o número de neurônios na camada de saída, com $0.5 \leq \alpha \leq 2$.

$$N_h = \alpha\sqrt{N_i X N_o} \tag{2}$$

Para o problema de classificação da face negativa considerando 17 pontos, $N_i = 34$ e $N_o = 1$. Assim, foram obtidos números no intervalo de 3 a 12. Foram geradas então combinações entre os números dentro desse intervalo. Essas combinações representam a quantidade de neurônios distribuídos nas camadas ocultas. Dessa forma, foram obtidas 75 redes e, expandindo-se a sugestão, foram obtidas mais 35 redes com no máximo 24 neurônios distribuídos pelas camadas ocultas, totalizando assim 110 redes, sendo 10 redes com uma camada oculta e 100 redes com duas.

4.3. Plataforma e bibliotecas utilizadas

A plataforma utilizada para desenvolver as redes propostas foi a linguagem de programação Python. A biblioteca Pandas, uma biblioteca de código aberto do Python, fornece estruturas de dados de alto desempenho e ferramentas de análise de dados e foi utilizada para auxiliar a manipulação do *dataset* [Pandas 2017]. Para os procedimentos matemáticos foi usado o NumPy, um pacote fundamental para computação científica com Python [NumPy 2017]. Já para o desenvolvimento das redes neurais foi utilizado a biblioteca scikit-learn, uma ferramenta simples e eficiente para mineração e análise de dados [Scikit-learn 2017].

5. Resultados e Discussão

O conjunto de dados foi particionado em 70% para conjunto de treinamento e 30% para o conjunto de teste, as redes foram treinadas 200 vezes utilizando como critério de parada

Camadas ocultas	*F-score*	*Precision*	*Recall*
9	0.778	0.782	0.792
10	0.789	0.791	0.803
11	0.785	0.804	0.780
12	0.784	0.797	0.788
(12,12)	0.778	0.789	0.782
(10,12)	0.776	0.796	0.771
(12,10)	0.775	0.786	0.780
(12,11)	0.774	0.785	0.772
(11,11)	0.773	0.794	0.761
(11,12)	0.771	0.783	0.775

Tabela 3: Arquiteturas das 10 redes com melhores resultados e suas métricas de desempenho.

o *early stopping criteria* e foram selecionadas as 10 melhores redes baseadas no *F-score*, como verifica-se na tabela 3.

Todas as 10 redes apresentam *F-score* em torno de 77%. Isso mostra uma maior capacidade das redes em detectar expressões negativas e um baixo desempenho em reconhecer expressões não negativas em decorrência destas representarem outras expressões faciais. Dentre essas redes, a que obteve melhor *F-score* foi a que possui uma camada oculta com 10 neurônios.

A fim de analisar os resultados obtidos, foram considerados os resultados do trabalho [de Almeida Freitas et al. 2014b] para comparar o desempenho das redes e verificou-se que todas apresentaram resultados superiores. Além disso, foi gerado um *F-score* para todas as outras expressões faciais a partir da configuração de rede com duas camadas ocultas com 10 e 12 neurônios, respectivamente. Para as expressões interrogativas binárias, condicionais e relativas foram obtidos melhores resultados, em contraste aos da literatura [de Almeida Freitas et al. 2014b], uma vez que os valores dos pontos para estas expressões possuem maior complexidade, como pode ser visto na Figura 2.

6. Considerações Finais

Este artigo realizou um estudo sobre as Expressões de Gramática Facial utilizadas no reconhecimento de linguagem de sinais brasileira, empregando uma técnica de Aprendizagem de Máquina. Nos experimentos realizados, utilizou-se os nove tipos de expressões faciais gramaticais com ênfase em expressões negativas, e o problema de identificação foi modelado por meio de um conjuntos de tarefas de classificação binária.

A partir da utilização de Redes Neurais, os resultados obtidos foram comparados com resultados disponíveis na literatura. Com isso, o método proposto obteve um *F-score* de, aproximadamente, 77% na detecção de expressões faciais negativas, um aumento de mais de 20% em relação ao [de Almeida Freitas et al. 2014b], o que torna a detecção dessas expressões uma alternativa viável para trabalhos que possam necessitar desse tipo de análise.

Por fim, com um bom resultado adquirido nos experimentos, empregou-se a

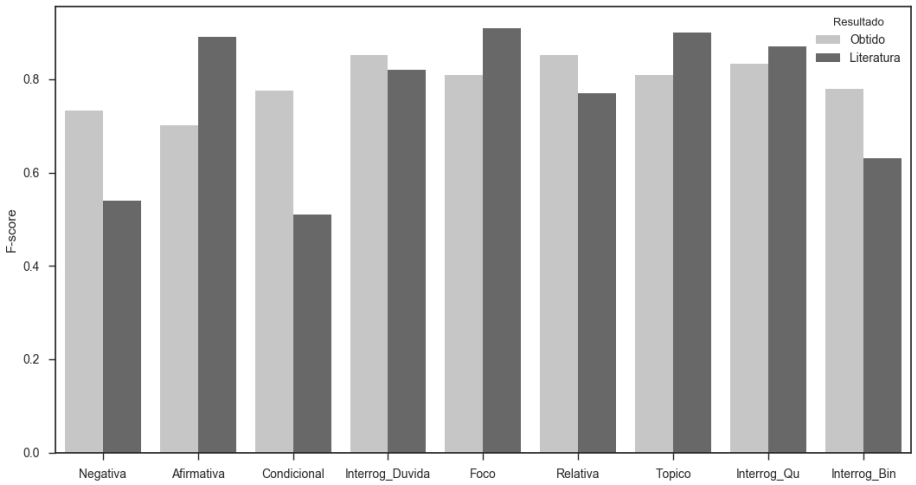

Figura 2: Comparativo de F-score entre os resultados obtidos e os da literatura.

mesma configuração de rede neural utilizada na detecção de expressões negativas, para as demais expressões faciais. Em mais da metade dos diferentes tipos de expressões, também foram obtidos melhores resultados em comparação com os resultados de [de Almeida Freitas et al. 2014b].

Como trabalhos futuros, pretende-se analisar mais pontos da face, acrescentado-os aos 17 utilizados, e relacionar a expressão facial com o movimento de mãos. Além disso, será considerado a utilização de mais usuários para a formação de um conjuntos de dados mais robusto, atendendo os diferente tipos de expressões que uma pessoas pode produz.

Referências

Arroteia, J. (2005). O papel da marcação não-manual nas sentenças negativas em língua de sinais brasileira (lsb).

de Almeida Freitas, F., Barbosa, F. V., and Peres, S. M. (2014a). Grammatical facial expressions data set. `https://archive.ics.uci.edu/ml/datasets/Grammatical+Facial+Expressions`. Acessado em 26 de fevereiro de 2018.

de Almeida Freitas, F., Barbosa, F. V., and Peres, S. M. (2014b). Grammatical facial expressions recognition with machine learning.

de Fátima Brecailo, S. (2012). Expressão facial e corporal na comunicação em Libras. `http://www.imap.curitiba.pr.gov.br/wp-content/uploads/2014/03/apostila_curso_expressao_corporal%20(1).pdf`.

Faceli, K., Lorena, A. C., Gama, J., and de Carvalho, A. (2011). *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. LTC.

Fernando de Almeida, Felipe Barbosa, S. M. (”2014”). Grammatical facial expressions data set. [Online; acessado em 20-Maio-2017].

Flores, E. M., Barbosa, J. L. V., and Rigo, S. J. (2012). Um estudo de técnicas aplicadas ao reconhecimento da língua de sinais: novas possibilidades de inclusão digital.

Lima, P., Filho, H., Lima, R., Oliveira, R., and Neto, A. (2011). Classificação de qos em conteúdo multimídia para rede vpn utilizando rede neural multilayer perceptron.

NumPy (2017). Numpy. http://www.numpy.org/.

Pandas (2017). Python data analysis library. http://pandas.pydata.org/.

Quadros, R. M. and Karnopp, L. B. (2009). *Língua de sinais brasileira: estudos lingüísticos*. Artmed Editora.

Richter, T., da Silva, E., and Gonzaga, A. (2007). Usando mlp para filtrar imagens.

Scikit-learn (2017). Scikit-learn - machine learning in python. http://scikit-learn.org/stable/.

Senger, L., Marcolino, E., Mello, R., and Souza, M. (2007). Javaart: Javaart: uma ferramenta computacional para aprendizado de máquina através da família de redes neurais artificiais art.

Soares, D. and Teive, R. (2015). Previsão de cheias do rio itajaí-açu utilizando redes neurais artificiais.

Sousa, D. V. C. (2011). Um olhar sobre os aspectos linguísticos da língua brasileira de sinais. *Littera*, 1(2).

Souza, E. and Monteiro, J. A. (2009). Estudo sobre sistema de detecção de intrusão por anomalias: uma abordagem utilizando redes neurais.

Vellasco, M. M. B. R. (2007). Redes neurais artificiais.

Wagner, P., Madeo, R., Peres, S., and Lima, C. (2013). Segmentação de unidades gestuais com multilayer perceptrons.