NUSP: 10423932
LIZ MARIBEL HUANCAPAZA HILASACA

TEXTURA E COR NA CLASSIFICAÇÃO DE SON

I. Descrição do problema:

Comumente a manipulação dos dispositivos moveis é feito de forma dactilar, uma forma poco comum de manipular os dispositivos moveis pode ser feito por meio de comandos de voz. Alguns destes comandos podem ser: "Yes", "No", "Up", "Down", "Left", "Right", "On", "Off", "Stop", "Go", "Zero", "One", "Two", "Three", "Four", "Five", "Six", "Seven", "Eight", and "Nine". Dessta forma a competição *TensorFlow Speech Recognition Challenge*[1] tem como desafio criar um algoritmo para o reconhecimento da fala, especificamente o reconhecimento dos comandos de voz descritos anteriormente. O conjunto de dados[2] empregado na competição consta de 65,000 audios de 1 segundo de duração de 30 palabras grabadas de cientos de personas diferentes.

II. Solucão do problema:

A solução planteada visa principalmente a extração de carateristicas de cor e textura desde os espectrogramas criados a partir dos audios. A hipotesis que se tem é que as carateristicas de textura e cor presentes nos espectrogramas podem ajudar ao reconhecimento de comandos de voz. Assim, a solução é descrita nos passos a seguir:

1. **Criar o spectrograma**: para cada arquivo de audio é extraido o espectograma empregando o mel spectrogram em escala logaritmica.

2. **Adicionar cor ao espectrograma**: seguindo um determinado mapa de cores no sistema RGB são atribuidas cores às magnitudes do espectograma.

3. **Extrair carateristicas de textura**: dada uma imagem de espectograma é transformada em escalas de cinza, e dada uma matriz M de dimensões 256x256 onde serão armazenadas as ocorrências das intensidades entre os pixels pi(x, y) e pj(x+dx, y+dy). Após de preencher a matriz M, são calculadas as 6 medidas estadísticas de *Haralick*: *Maximum probability*, *Correlation*, *Contrast*, *Energy*, *Homogeneity* e *Entropy*. São empregadas 6 variações de deslocameto para pj, sendo Q(dx,dy)={(0,1), (0,3), (0,5), (1,0), (3,0), (5,0)}. O vetor resultante para a textura tem 36 dimensões.

4. **Extrair carateristicas de cor**: dada uma imagem de espectrograma, primeiro é reduzido o espaço de cor a 64 cores. São definidos 2 vetores V1 e V2 de 64 dimensões cada. No vetor V2 são armazenados os pixeis com a recorrência de cor na mesma região. No vetor V1 são armazenados os pixeis com a recorrência de cores localizados nas bordas da imagem. Para computar as caraterísticas de cor, os pixeis são percorridos em relação a sua vizinhança, se o pixel p é igual a seus 4 vizinhos é considerado com pixel interior sendo armazenado como um incremento em V2[p], caso contrario o pixel é armazenado como borda no incrementeo para V1[p]. O vetor resultante é a concatenação de V1 e V2 que possui 128 dimensões.

5. **Classificação supervisionada**: neste passo são realizadas as tarefas de treinamento e teste com as carateristicas extraidas de cor e textura. Os resultados são avaliados com a accuracia de classificação. Assim para o aprendizagem são empregados os classificadores: *Random Forest (RF)*, *Gaussian Naive Bayes (GNB)*, *K-Nearest Neighbors (KNN)* e *Super Vector Machine (SVM).*

1 https://www.kaggle.com/c/tensorflow-speech-recognition-challenge
2 https://www.kaggle.com/c/tensorflow-speech-recognition-challenge/data

III. Resultados e discuções:

Para gerar os resultados foram selecionados 2 classes do conjunto dataset, sendo as classes “go” e “stop”, com 506 instancias entre as 2 classes. A Figura 1 ilustra os 2 tipos de mapas de cores empregados para colorir os espectogramas. A Figura 2 mostra um exemplo do resultado dos espectogramas com os 2 mapas de cor.

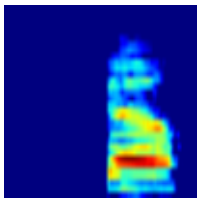
As Tabelas 1 e 2, assim como as Figuras 3 e 4 apresentam os resultados de accuracia obtidos na classificação com as caraterísticas de textura e cor extraidas desde os espectogramas. Nos resultados obtidos se pode apreciar a superioridade do descritor de textura GLCM ao ser comparado com o descritor de cor BIC, esta superioridade es determinada com a melhor acuracia e ao possuir o menor número de carateristicas. Também pode ser notar um clara vantagem em colorir os espectogramas com o mapa de cor Inferno (Fig. 1(b)).

Futuras extensões de este experimento poderiam proporcionar informação ao evaluar outros mapas de cor. Pode-se também empregar outros metodos para criar espectrogramas. É importante salientar que neste experimento soamente formam empregadas 2 classes do dataset. Para incluir as demais classes pode ser incluido também carateristicas presentes no sinal que representa ao son.

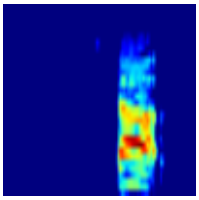
(a)

(b)

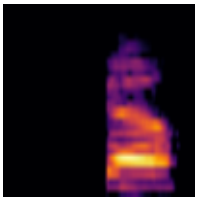
Figura 1: mapas de core empregados para gerar os espectrogramas: (a) Jit, (b) Inferno



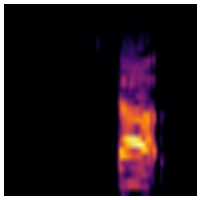
(a)



(b)



(c)



(d)

Figura 2: visualização dos espectogramas : (a) e (c) para o son “go”, (b) e (d) para o son “stop”

Tabla 1: resultados de acuracia da classificação (10-k-flod) empregando o mapa de cor Jet

Classificador	GLCM (36)	BIC (128)	GLCM+BIC (36+128)
RF	**84.18**	69.36	81.22
GNB	**63.64**	55.98	57.33
KNN	**73.71**	67.95	73.11
SVM	78.06	**76.47**	**83.80**

Tabla 2: resultados de acuracia da classificação (10-k-flod) empregando o mapa de cor Inferno

Classificador	GLCM (36)	BIC (128)	GLCM+BIC (36+128)
RF	**85.60**	66.24	**86.17**
GNB	66.19	53.14	**70.50**
KNN	**73.53**	58.89	69.56
SVM	82.61	**71.35**	**84.77**

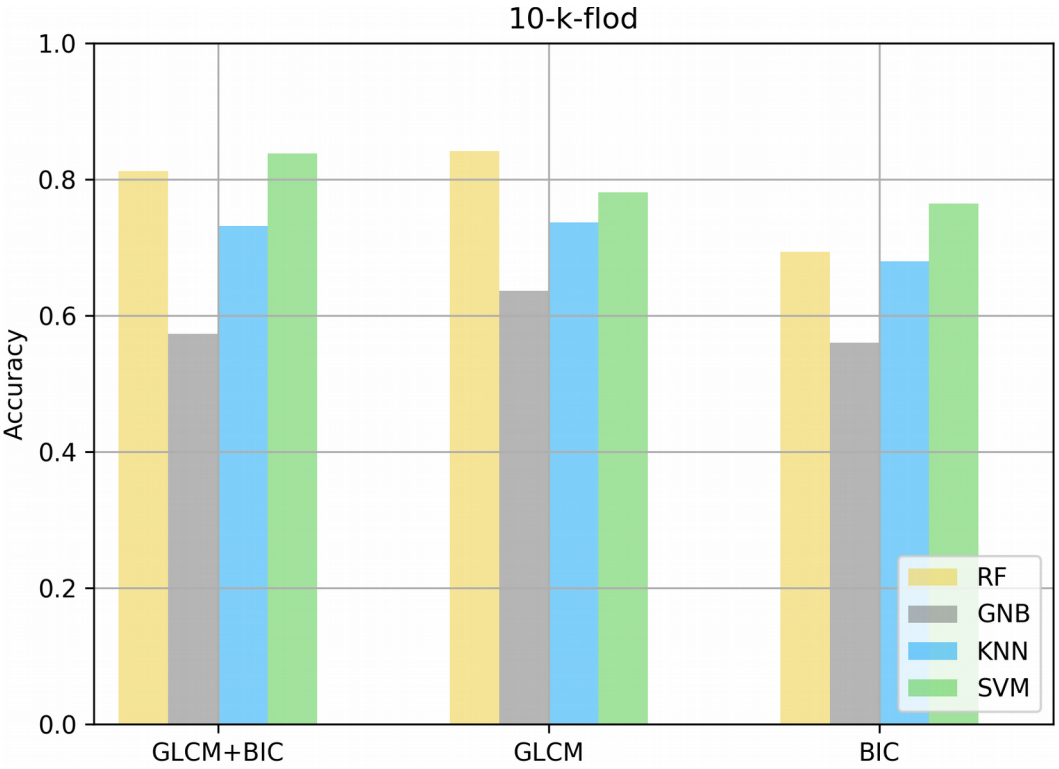


Figura 3: resultados da acuracia da classificação para as classe "go" e "stop" com o mapa de cor Jet

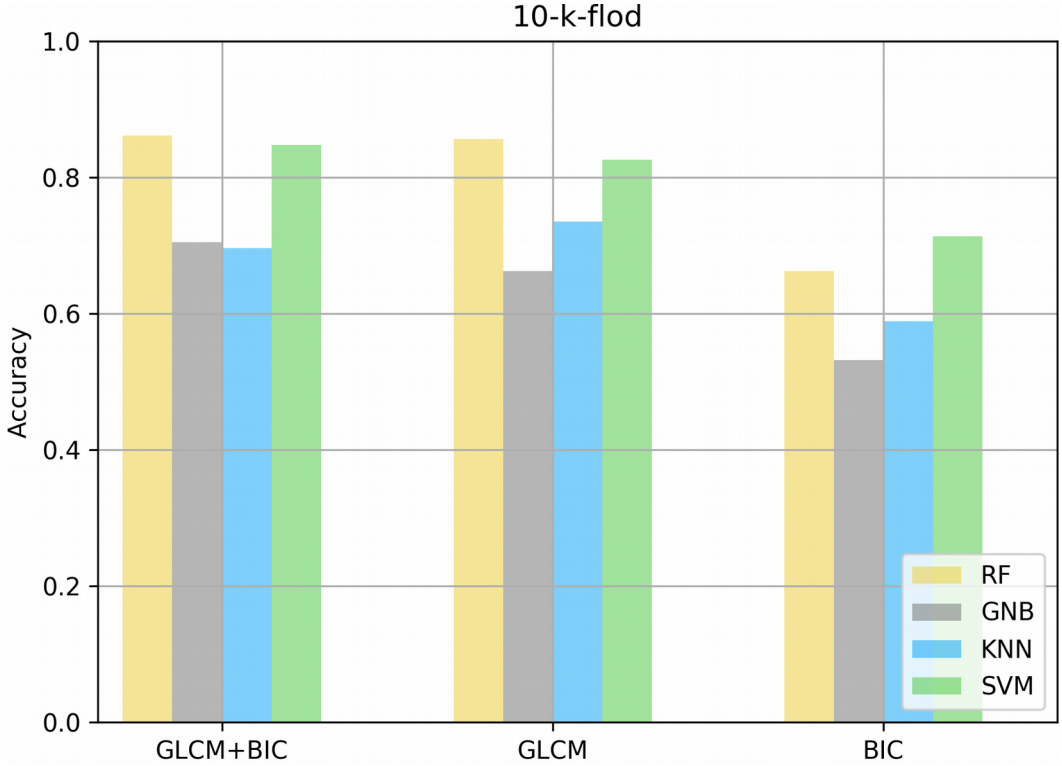


Figura 4: resultados da acuracia da classificação para as classe “go” e “stop” com o mapa de cor Inferno