

TEXTURA E COR NA CLASSIFICAÇÃO DE SON

Liz Maribel Huancapaza Hilaraca
(NUSP: 10423932)

I. Descrição do problema:

Comumente a manipulação dos dispositivos moveis é feito com a interação dos dedos. Uma forma pouco 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". Desta forma a competição TensorFlow Speech Recognition Challenge¹ tem como desafio criar um algoritmo para o reconhecimento da fala, especificamente o reconhecimento dos comandos de voz descritos anteriormente. O conjunto de dados² empregado na competição consta de 65,000 áudios de 1 segundo de duração de 30 palavras gravadas de centos de personas diferentes.

II. Solução do problema:

A solução planteada visa principalmente a extração de caraterísticas de cor e textura desde os espectrogramas criados a partir dos áudios. A hipóteses que se tem é que as caraterísticas 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 espectrograma:** para cada arquivo de áudio é extraído o espectrograma empregando o método log mel spectrogram.
2. **Adicionar cor ao espectrograma:** seguindo um determinado mapa de cores no sistema RGB são atribuídas cores às magnitudes do espectrograma.
3. **Extrair caraterísticas de textura:** dada uma imagem de espectrograma é transformada em escalas de cinza, e dada uma matriz M de dimensões 256×256 onde serão armazenadas as ocorrências das intensidades entre os pixels $p_i(x, y)$ e $p_j(x+dx, y+dy)$. 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 deslocamento para p_j , 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 caraterísticas 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 pixels com a recorrência de cor na mesma região. No vetor $V1$ são armazenados os pixels com a recorrência de cores localizados nas bordas da imagem. Para computar as caraterísticas de cor, os pixels são percorridos em relação a sua vizinhança, se o pixel p é igual a seus 4 vizinhos é considerado como pixel interior sendo armazenado como um incremento em $V2[p]$, caso contrario o pixel é armazenado como borda no incremento 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 caraterísticas extraídas de cor e textura. Os resultados são avaliados com a acurácia de classificação. Assim para a aprendizagem são empregados os classificadores: Random Forest (RF), Gaussian Naive Bayes (GNB), K-Nearest Neighbors (KNN) e Super Vector Machine (SVM).

III. Resultados e discussões:

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 espectrogramas. A Figura 2 mostra um exemplo do resultado dos espectrogramas com os 2 mapas de cor.

As Tabelas 1 e 2, assim como as Figuras 3 e 4 apresentam os resultados de acurácia obtidos na classificação com as características de textura e cor extraídas desde os espectrogramas. 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 acurácia e ao possuir o menor número de caraterísticas. Também pode ser notar uma clara vantagem em colorir os espectrogramas com o mapa de cor Inferno (Fig. 1(b)).

Futuras extensões de este experimento poderiam proporcionar informação ao avaliar outros mapas de cor somente para as regiões que contem o son. Pode-se também empregar outros métodos para criar espectrogramas. É importante salientar que neste experimento somente formam empregadas 2 classes do dataset. Para incluir as demais classes pode ser incluído também caraterísticas presentes no sinal que representa ao son.

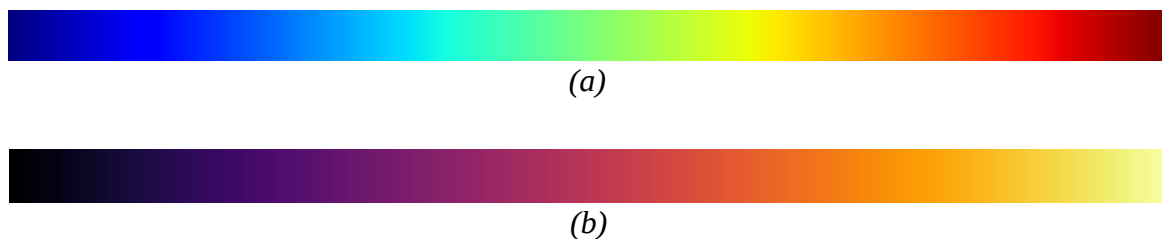


Figura 1: mapas de core empregados para gerar os espectrogramas: (a) Jet, (b) Inferno

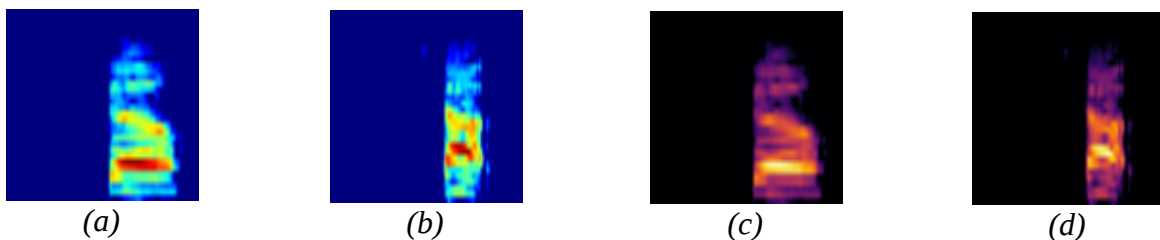


Figura 2: visualização dos espectrogramas : (a) e (c) para o son “go”, (b) e (d) para o son “stop”

Tabela 1: resultados de acurácia 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

Tabela 2: resultados de acurácia 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

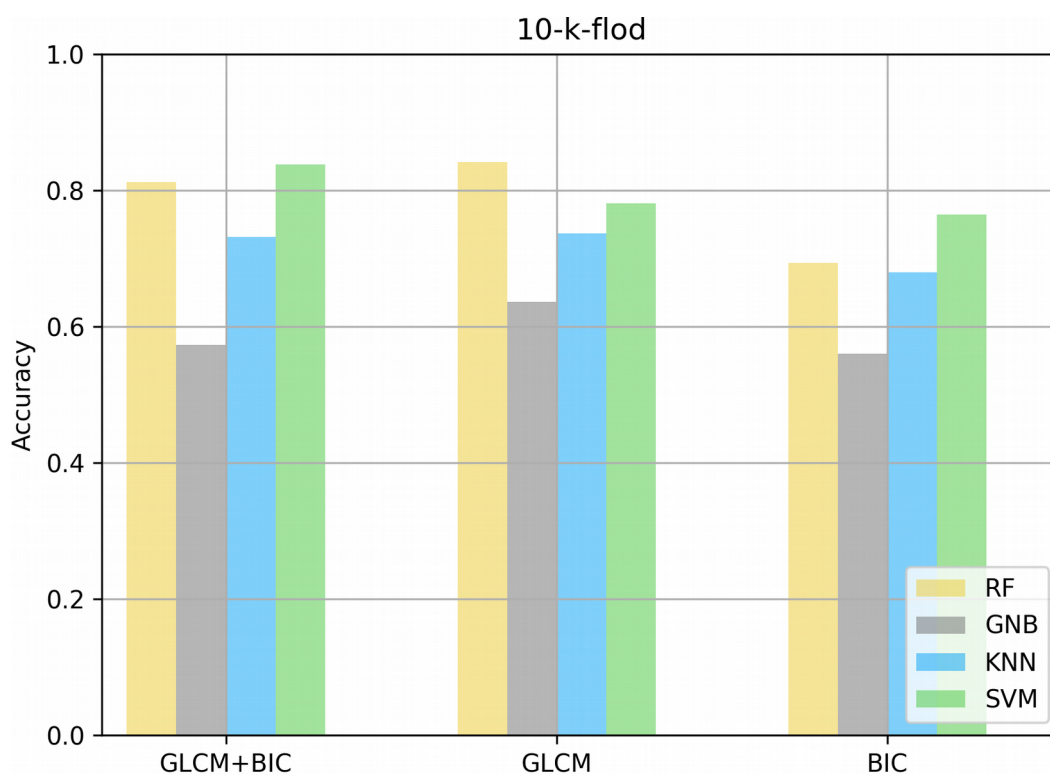


Figura 3: resultados da acuracia da classificação para as classes "go" e "stop" com o mapa de cor Jet

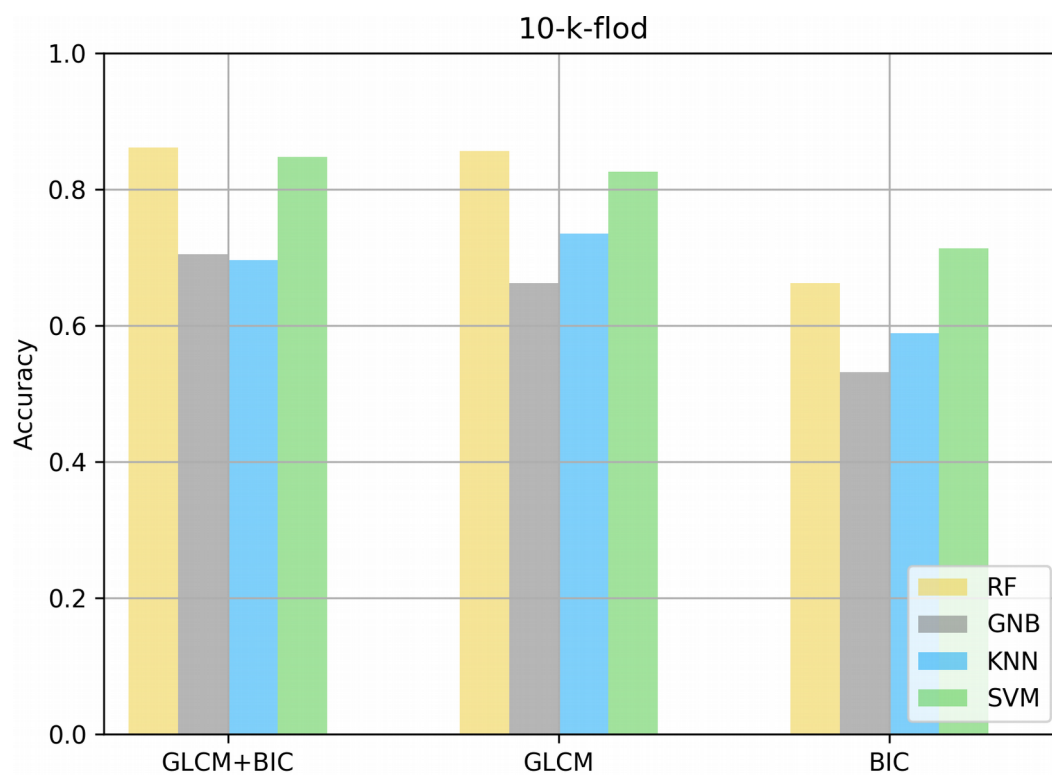


Figura 4: resultados da acuracia da classificação para as classes “go” e “stop” com o mapa de cor Inferno