CRIAÇÃO DE BASE COM AMOSTRAS DE FALA EM LÍNGUA JAPONESA

Edson Matheus Alexandre Cizeski (PIBIC/CNPq/FA/UEM), Yandre Maldonado e Gomes da Costa (Orientador), e-mail: yandre@din.uem.br., Flávio Rogério Uber (Coorientador).

Universidade Estadual de Maringá / Centro de Tecnologia / Maringá, PR.

Ciências Exatas e da Terra / Ciência da Computação

Palavras-chave: base de dados, reconhecimento de padrões, japonês

Resumo:

Este projeto visa expandir a Brazilian Speech Dataset (BSD), base de dados já criada pelo grupo de pesquisa proponente deste projeto, de forma que ela passe a ser uma base multilíngue, incorporando amostras de fala em língua japonesa. O protocolo básico para coleta e organização das amostras seguiu, na medida do possível, o padrão já utilizado para as amostras em língua portuguesa. Foram colhidas amostras de áudio da leitura de 5 textos previamente selecionados para cada voluntário participante. Foi feito um esforço no sentido de que os voluntários escolhidos levassem à uma base com distribuição equilibrada entre gênero (masculino/feminino) e faixa etária, conforme já feito na BSD.

Introdução

A geração de recursos para apoio ao desenvolvimento de pesquisas em reconhecimento de padrões é bastante importante para viabilizar novas descobertas neste contexto de investigação científica. Um dos obstáculos mais conhecidos por pesquisadores da área é, em muitos casos, a dificuldade para se obter uma base de dados com características apropriadas para o desenvolvimento de trabalhos de pesquisa. Este projeto tem como objetivo colaborar com a expansão da base BSD, que em sua primeira versão conta apenas com amostras de áudio de fala em língua portuguesa. Além da expansão da base, trazendo amostras de fala da língua japonesa, também será possível com a realização deste projeto o desenvolvimento de alguns experimentos iniciais para a classificação de amostras de áudio a partir de subconjuntos de dados contidos nela, considerando a classificação por gênero e faixa etária.

Materiais e métodos

Sobre o método da realização da pesquisa, primeiro foram escolhidos cinco textos em língua japonesa para serem lidos pelos voluntários na hora da gravação. Uma vez escolhidos os textos, identificamos voluntários para colher amostras de fala em língua japonesa, gravadas em formato de áudio digital, levando-se em conta uma distribuição equitativa entre gêneros e faixas etárias, tal como feito no trabalho de Paulino et al. (2018) [1], em que foi elaborada uma base com amostras de fala em língua portuguesa.

Uma vez identificados os voluntários em quantidade suficiente, as amostras foram colhidas, utilizando um celular com microfone, e a base de dados foi criada e organizada de maneira apropriada para que pesquisadores da área de recuperação de informações multimídia possam utilizar no desenvolvimento de pesquisas ligadas a classificação de fala.

Adicionalmente, um protocolo experimental básico de classificação foi aplicado a base a fim de gerar resultados de baseline, que outros pesquisadores possam utilizar como um referencial para avaliar os resultados que eventualmente venham a obter.

Resultados e Discussão

Em relação aos resultados obtidos, o processo de identificação de voluntários foi mais difícil que o esperado, principalmente nas faixas etárias de 0 a 14 anos e na de mais de 55 anos. Mesmo assim, um total de 56 voluntários participou da pesquisa, e cinco amostras de áudio foram coletadas para cada participante, formando então uma base de 280 amostras distribuídas conforme dados apresentados na Tabela 1.

Tabela 1 - Quantidade de amostras por faixa etária

	de 0 a 14 anos	**de 15 a 24 anos**	**de 25 a 54 anos**	**mais de 55 anos**
Masculino	15	50	65	-
Feminino	30	55	50	15

Depois de organizada a base de dados, alguns testes iniciais foram realizados. O foco principal foi avaliar o desempenho de esquemas de classificação de áudio com o uso de imagens de espectrograma, uma representação visual do sinal do áudio muito utilizada para este fim desde 2011 [3]. Essas imagens foram geradas a partir dos áudios obtidos utilizando a ferramenta SoX(Sound Exchange), e com elas extraímos as features

necessárias para a classificação SVM, mais especificamente a feature Local Binary Pattern(LBP) [2], feita através do framework LIBSVM.

Os testes foram feitos utilizando k-folds cross validation, que consiste basicamente em uma divisão da base de dados em *k* subconjuntos de tamanho aproximadamente igual, e então são feitas *k* rodadas de classificação onde em cada uma (□ – *1*)/□ da quantidade de dados disponível é utilizada para o treino e *1*/□ é utilizado para o teste. O valor escolhido para k foi 5.

Os resultados da classificação de gênero e faixa etária estão descritos na Tabela 2, divididos entre os 5 folds obtidos através da técnica descrita acima, e a média correspondente:

Tabela 2 - Resultados da classificação utilizando espectrogramas

	0	**1**	**2**	**3**	**4**	**Média**
Gênero	83,92%	94,64%	96,42%	94,64%	100%	93,82%
Faixa Etária	58,92%	69,64%	73,21%	82,14%	71,42%	71,07%

Por fim, realizamos testes utilizando os próprios áudios a partir da extração da features de MFCC (Mel-frequency cepstrum coefficients), utilizando a ferramenta MIRTOOLBOX para tal. Da mesma forma, dividimos a base em 5 folds e testamos a classificação de gênero e faixa etária. Os resultados estão descritos a seguir:

Tabela 3 - Resultados da classificação utilizando MFCC

	0	**1**	**2**	**3**	**4**	**Média**
Gênero	62,5%	57,14%	80,35%	75%	60,7%	67,1%
Faixa Etária	58,9%	57%	80,35%	78,57%	64,28%	67,82%

Conclusões

Com a conclusão do projeto, conseguimos com sucesso cumprir o objetivo de expandir a base de dados, apesar das dificuldades encontradas na identificação de voluntários. Além disso, os resultados iniciais das classificações também atingiram um resultado satisfatório, revelando ainda que a classificação através das imagens de espectrograma obtiveram maiores taxas de acerto em comparação aos testes feitos utilizando características acústicas. Portanto, fica claro que os principais objetivos da pesquisa foram cumpridos e que futuros testes feitos com a nova base de

dados pode apresentar resultados bastante promissores no campo de reconhecimento de padrões.

Agradecimentos
Em primeiro lugar, agradecemos todos os voluntários que cederam seu tempo para participar da pesquisa e à Fundação Araucária pela bolsa concedida para o desenvolvimento deste trabalho. Também ressaltamos e agradecemos todo o trabalho dos professores do Instituto de Estudos Japoneses (IEJ) que contribuíram com a seleção dos textos. Em especial a Profa. Kiyomi Kimura Fugie, que deu apoio fundamental na divulgação do projeto para possíveis voluntários, e por todo esforço e ajuda que oferecida para que a pesquisa pudesse ocorrer. Por fim, agradecer aos professores da Escola de Língua Japonesa da ACEMA (Associação Cultural e Esportiva de Maringá), que também apoiaram a coleta de amostras.

Referências
[1] PAULINO, M. A. D.; SVAIGEN, A. R.; COSTA, Y. M. G.; AYLON, L. B. R.; BRITTO JR., A. S.; OLIVEIRA, L. E. S. **A Brazilian speech database**, 30th International Conference on Tools with Artificial Intelligence, Volos, Greece, 2018.

[2] COSTA, Y. M. G.; Oliveira, L.S. ; Koerich, A.L.; GOUYON, F.; Martins, J.G. **Music genre classification using LBP textural features.** Signal Processing (Print) , v. 92, p. 2723-2737, 2012.

[3] COSTA, Y. M. G.; OLIVEIRA, L. E. S.; KOERICH, A. L.; GOUYON, F. **Music Genre Recognition using Spectrograms.** In: 18th International Conference on Systems, Signals and Image Processing, 2011, Sarajevo. Proceedings IWSSIP 2011. Sarajevo: V-graf d.o.o Sarajevo, 2011. v. 1. p. 151-154.