



SEGMENTAÇÃO AUTOMÁTICA DO SINAL DE ÁUDIO PARA TAREFAS DE CLASSIFICAÇÃO

Juliano César Chagas Tavares (PIBIC/CNPq/FA/Uem), Yandre Maldonado e Gomes da Costa (Orientador), e-mail: ra87940@uem.br

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

Área: Ciências Exatas e da Terra / Subárea: Ciência da Computação

Palavras-chave: segmentação dinâmica, classificação automática, classificação de pássaros

Resumo:

A pesquisa teve como objetivo descobrir técnicas que possam ser aplicadas à segmentação automática de sinais de áudio para tarefas de classificação a fim de poupar esforço humano para o tratamento dos sinais, porém sem perder a qualidade dos áudios e as informações neles presentes. Para esse trabalho, o foco foi dado para os sons dos pássaros. Foram utilizadas amostras de áudio de pássaros disponibilizados pelo site Xeno-Canto¹. Para realizar a segmentação, foram empregadas técnicas de remoção de ruído característico das amostras e de descoberta de regiões de interesse utilizando o centroide espectral e a energia do sinal. Os resultados obtidos foram satisfatórios, chegando a taxas de acerto próximas às alcançadas utilizando a segmentação manual, que demanda um trabalho mais demorado, porém mais preciso. Com o estudo foi possível observar também que é possível melhorar ainda mais os resultados obtidos melhorando as técnicas de remoção de ruído aplicadas nesse estudo.

Introdução

Com o avanço das tecnologias, desenvolvimento de sensores, gravadores de áudio cada vez menores e mais baratos, celulares, as possibilidades de identificação do pássaro a partir da sua vocalização ficaram ainda mais atraentes. Surgem então estudos que se aprofundam

¹ www.xeno-canto.org





ainda mais no trabalho de classificação de espécies a partir de técnicas de processamento de sinais juntamente com técnicas de aprendizado de máquina [1].

Porém, nem sempre a coleta de amostras de áudio é feita com equipamentos profissionais ou em estúdios. Em geral o maior volume de dados é coletado de forma amadora e ao ar livre, o que traz características às amostras que podem dificultar a identificação das espécies. De acordo com Bardelo et al. [2] e Agranat [3], em ambientes reais há problemas como o ruído causado pelo vento, assim como também a sobreposição de sinais de áudio, que acabam por afetar mais de cinquenta por cento das gravações.

Materiais e métodos

Para a execução do trabalho foi utilizada a linguagem MATLAB, amplamente utilizada na área de processamento de sinais. Foi iniciado o estudo acerca de processamento de sinais e baseado no algoritmo de [5] foi iniciada a construção do segmentador. Uma das ferramentas utilizadas foi o software Sound eXchange (SoX), versão 14.4.1, disponível gratuitamente em <http://sox.sourceforge.net/>. As etapas que compreendem o processo de segmentação são: identificação do ruído padrão de cada espécie, remoção do ruído de cada amostra usando SoX, aplicação da segmentação a partir da busca por regiões de interesse.

Para identificar o ruído de cada uma das amostras de pássaros, foi implementado um algoritmo que se baseia em duas características que podem ser encontradas em um sinal, o Centroide Espectral (*Spectral Centroid*) e a Energia do Sinal (*Signal Energy*). Além de serem funções simples de implementar, tendo um custo computacional justo, trazem informações relevantes à aplicação. Com isso é possível obter dados que diferenciam o som do animal e o ruído do ambiente. Sua descrição é dada abaixo:

- Energia do Sinal: Quando o nível de ruído não é muito alto, a energia da região de interesse será maior que a energia das demais regiões;
- Centroide Espectral: Se sons e ruídos estiverem em uma frequência mais baixa do que a região de interesse, o Centroide Espectral será maior do que nas demais regiões;

Na segunda etapa as amostras de ruído geradas na primeira etapa são concatenadas, gerando um ruído padrão da espécie, usando o SoX. Com





esse software também é feita a remoção do ruído. Para tal é necessário informar a amostra a ser tratada, o ruído padrão a ser removido e um parâmetro que indica a intensidade da limpeza. Ao final da limpeza o software produz uma amostra de áudio contendo apenas o que ele considerou como não compatível com o ruído padrão informado.

Na última etapa é aplicado o mesmo algoritmo explicado na primeira etapa, com a diferença de que ele irá buscar e salvar a região de interesse, e não mais o ruído como anteriormente. A remoção de ruído é importante para que fique mais evidente na amostra quais são as reais regiões de interesse, já que em um ambiente externo é possível que haja ruídos com uma intensidade forte, podendo ser confundido com o pássaro. Com isso é obtido um áudio final com uma redução considerável de ruído e de duração.

Resultados e Discussão

Para a coleta dos dados foi realizado um estudo na etapa de remoção de ruído, já que essa se trata de uma etapa fundamental para todo o processo. Os melhores resultados obtidos no estudo tiveram a remoção de ruídos aplicada em duas etapas. Uma primeira remoção de ruídos utilizando a intensidade 0,5, em que são removidos os ruídos mais grosseiros da amostra e uma segunda filtragem utilizando a intensidade 0,21 em que são removidos ruídos mais sutis restantes, valores ajustados empiricamente.

Na coleta dos dados foi utilizada a classificação utilizando espectrogramas como proposto em [4]. Para a extração de características foi usado o *Local Binary Pattern* (LBP) e para a classificação foi utilizada a *LibSVM* v3.2, uma implementação de *Support Vector Machine* (SVM).

A sub-base utilizada no trabalho foi composta por 47 espécies. O critério usado para selecionar as espécies foi a quantidade de amostras. Nesse trabalho todas as 47 espécies possuem 10 amostras, portanto a base final é composta por 470 amostras diferentes. Foram realizados experimentos com seis variações:

Tabela 1

	Original	Original Sem Ruído	Manual	Manual Sem Ruído	Automática	Automática Sem Ruído
Taxa de Acerto Total	22,13%	23,40%	30,64%	30,85%	25,74%	28,30%





Conclusões

As taxas de acerto com a base sem nenhum tratamento e a base já tratada são visivelmente diferentes. No caso a maior taxa de acerto pertence à base com a segmentação manual e com remoção de ruídos, o que reforça a ideia de que tratar o ruído é importante para a posterior classificação do som do animal. Porém a segmentação proposta nesse trabalho não ficou atrás, atingindo taxas de acerto próximas à melhor taxa obtidas. Uma possível continuação para esse trabalho poderia ser dada a partir de uma busca de métodos mais eficientes de encontrar o ruído característico da espécie e removê-lo.

Agradecimentos

Ao programa CNPq/FA/PIBIC pelo financiamento do projeto de pesquisa, à UEM pela concessão da bolsa de Iniciação Científica, ao professor Yandre Maldonado pelo apoio e confiança dados e ao Anderson S. Oyama que deu apoio fundamental em diversas etapas do projeto.

Referências

- [1] LOPES, M.; GIOPPO, L.; HIGUSHI, T.; KAESTNER, C.; SILLA, C.; KOERICH, A.; Automatic Bird Species Identification for Large Number of Species. In: IEEE INTERNATIONAL SYMPOSIUM ON MULTIMEDIA (ISM), 2011. p. 117-112.
- [2] BARDELLI, R.; WOLFF, D.; KURTH, F.; KOCH, M.; TAUCHERT, K. H.; FROMMOLT, K. H.; Detecting Bird sounds in a complex acoustic environment and application to bioacoustic monitoring. **Pattern Recognition Letters**, v. 31, n. 12, p. 1524-1534, 2010.
- [3] AGRANAT, I.; Automatically indentifying animal species from their vocalizations. In: FIFTH INTERNATIONAL CONFERENCE ON BIO-ACOUSTICS, Holywell Park, 2009.
- [4] COSTA, Y. M. G.; OLIVEIRA, L. E. S.; KOERICH, A. L.; GOUYON, F. Music genre recognition using spectrograms. In: INTERNATIONAL CONFERENCE ON SYSTEMS, SIGNALS AND IMAGEM PROCESSING, Sarajevo, Bosnia and Herzegovina, 2011.
- [5] GIANNAKOPOULOS, T.; A method for silence removal and segmentation of speech signals, implemented in matlab. University of Athens, Athens, 2009.

