

Table of Contents

- Chapter 1. Introduction
 - 1.1 Natural Language Understanding
 - 1.2 Keyword Spotting
- Chapter 2. Project Objectives
 - 2.1 Problem Specification
 - 2.2 Functional Requirements
 - 2.3 Non-functional Requirements
- Chapter 3. Bibliographic Research
 - 3.1 Automatic Speech Recognition
 - 3.2 Unconventional Approaches
 - 3.2.1 Audio Feature Extraction
 - 3.2.2 Feature Comparison
 - 3.2.1 Keyword Prediction
- Chapter 4. Analysis and Theoretical Foundation
 - 4.1 Conceptual View
 - 4.2 Data Pre-Processing
 - 4.2.1 Mel-Frequency Cepstral Coefficients
 - 4.2.2 Gaussian Mixture Model
 - 4.2.3 Dynamic Time Warping
 - 4.3 Keyword Classification
 - 4.3.1 Convolutional Neural Network Architecture
 - 4.4 Dataset
 - 4.4.1 General Aspects
 - 4.4.2 Data set partitioning
 - 4.5 Feature Selection
- Chapter 5. Detailed Design and Implementation
 - 5.1 System Architecture
 - 5.2 Feature Extraction
 - 5.3 Image Processing Method
 - 5.4 Training
- Chapter 6. Testing and Validation
 - 6.1 Implementation Validation on Larger Dataset

- 6.2 Implementation Validation on Reduced Dataset
- 6.3 Running Time Evaluation
- Chapter 7. User's Manual
 - 7.1 System Setup
 - 7.2 System Usage
- Chapter 8. Conclusions
 - 8.1 Contributions and Achievements
 - 8.2 Result Analysis
 - 8.3 Further Work
- Bibliography

References

- [1]** A. Graves, S. Fernandez, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks", In ICML, 2006.
- [2]** Liu, Bing and Ian Lane. "Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling." INTERSPEECH, 2016
- [3]** Shankar, Ravi et al. "Spoken Keyword Detection Using Joint DTW-CNN." Interspeech, 2018
- [4]** Gokhan Tur, Dilek Hakkani-Tur, Larry Heck, and Sarangarajan Parthasarathy, "Sentence simplification for spoken language understanding", In ICASSP, IEEE, 2011.
- [5]** Xia, Congying et al. "Zero-shot User Intent Detection via Capsule Neural Networks." EMNLP, 2018
- [6]** P. Xu and R. Sarikaya, "Convolutional neural network based triangular CRF for joint intent detection and slot filling," in Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on. IEEE, 2013.
- [7]** Yaodong Zhang and James Glass, "Unsupervised spoken keyword spotting via segmental DTW on Gaussian posteriorgrams", In Proceedings of ASRU, pages 398--403, 2009
- [8]** Zhang, Chenwei et al. "Joint Slot Filling and Intent Detection via Capsule Neural Networks." CoRR abs/1812.09471, 2018