Chapter 15
Similarity Measure Optimization for Target Detection:
A Case Study for Detection of Keywords in Telephone Conversations

Batuhan Gundogdu
National Defence University, Turkey

Murat Saraclar
Bogazici University, Turkey

ABSTRACT

Today, in the era of data and computing, fast and reliable retrieval of information has become of great importance for security and military applications, and continues to be such, as the amount available digital data increases every second. While the search and retrieval of text data has produced mature products and are today being used in search engines everyday by everyone, the retrieval of spoken content still remains a young research, especially for low resource languages where the available data is scarce to train reliable speech recognition systems. This chapter provides a thorough introduction of a speech retrieval task called "keyword search" and presents a novel similarity measure optimization-based approach. The case study was experimented on telephone conversations in three different languages and thousands of keywords randomly selected from each language were searched in the document. The experiments show that the technique introduced in this chapter offers a new methodology to handle the terms that does not even exist in the vocabulary of the speech recognition systems.

INTRODUCTION

The digitalization of data and the ever-increasing storage capacity has brought people closer to a life like the fictional world in Jorge Luis Borges' *Library of Babel* (Borges, 1998), —an infinite library *whose books contain every possible string of letters and, therefore, somewhere an explanation of why the library exists and how to use it.* Though, it was highly doubted that the librarians would ever find

that book amid the miles of nonsense. The situation the society faces today is somewhat similar: When it comes to retrieval of data, and especially the spoken data, the conundrum appears to be the "*Paradox of Abundance*" defined in Haidt (2006): *Quantity undermines the quality of our engagement*. Hence, it is vital now, in the Internet era; to efficiently and correctly retrieve the data of interest to make our lives easier as evidenced by the well-developed text retrieval technology, which is undoubtedly one of the biggest life-changing innovations of the past century. On the other hand, speech retrieval technology is still nascent even though speech is definitely one of the most appealing forms of information content. The retrieval of speech is therefore not only attractive but also necessary for users to efficiently browse across the vast quantities of multimedia content.

Detection of certain words in telephone conversations is also a very important task for security reasons and fight against terrorism. Such a task may require a dynamic list of 'suspicious' keywords, so that there no longer exists a necessity for developing the retrieval system from the beginning once new keywords become of interest.

This chapter focuses on a speech retrieval task called spoken term detection, also known as keyword search (KWS). In KWS, the user provides a query in text form consisting of one or more words. This query is searched in an un-transcribed audio archive and the locations in the audio where the query exits are returned to the user in an ordered list, with respect to a similarity score the KWS system attributes to them. The text counterpart of the task that was just defined is a very common one in everyday life. Nowadays, people rarely spend a day without searching something in Google©. Text browsing also comes in handy and used in text editors with very well-known short-cut button configurations. Despite this wide usage of text retrieval acknowledged by almost every computer user, searching for speech still remains a young research. Searching for audio-visual content in multimedia archives, such as YouTube, only provides videos whose title or meta-data has high similarity scores to the text query. The actual locations where the terms of interest are uttered are not available upon such a search. In KWS, on the other hand, the audio documents are returned to the user along with the time stamps where the query occurs. Such a system maybe desired in situations where the user is interested in finding locations of their queries in lecture videos, broadcast news reports, radio communications, meeting recordings. Furthermore, detecting and locating the utterance of certain keywords in telephone conversations may be very important for security applications for prevention and/or investigation of criminal or terrorist activities. System diagram demonstrating the KWS task is given in Figure 1.

Since the text retrieval task is now a mature product and has been providing reliable results, the first practiced approach to KWS was using an automatic speech recognition (ASR) system on the audio archive and converting the task into a text retrieval task. Once the spoken document is converted into a string of words by the ASR system, the task is now to finding text in text, which is not so hard. However, only depending on the ASR system output limits the retrieval system's performance to the accuracy and the vocabulary size of the ASR systems being used. Furthermore, when circumstances are harder as in the situations when the recording conditions are very noisy and there is little or no available data for the language of interest, which is referred to as a "*low resource language*" in the literature.

This chapter aims to address the problem of KWS especially for low resource languages by presenting the methodologies adopted in literature and introduces a novel approach that conducts a similarity measure optimization and a dynamic search. In the next section the definitions are presented and the conventional techniques are introduced. The problem brought by scarcity of resources is addressed in a thorough manner and the solution methodology is described. The validity of the approach is presented

26 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/similarity-measure-optimization-for-target-detection/209812

Related Content

A Fuzzy MCDM Framework for Weapon Systems Selection
Melis Almula Karadayi, Yeliz Ekinciand Hakan Tozan (2019). *Operations Research for Military Organizations (pp. 185-204).*
www.irma-international.org/chapter/a-fuzzy-mcdm-framework-for-weapon-systems-selection/209806

JomMachineLearning: Bringing Artwork Nearer With DesignLab
K.M. Liow, P.S. Joseph Ngand H.C. Eaw (2021). *International Journal of Business Strategy and Automation (pp. 54-71).*
www.irma-international.org/article/jommachinelearning/276457

Optimization of NAS Lemoore Scheduling to Support a Growing Aircraft Population
Manuel Rosasand Emily Craparo (2019). *Operations Research for Military Organizations (pp. 268-312).*
www.irma-international.org/chapter/optimization-of-nas-lemoore-scheduling-to-support-a-growing-aircraft-population/209809

Pricing Strategies for Companies During the COVID-19 Pandemic
Lovely Chopra, Rohan Shareshtha Vermaand Pratap Chandra Mandal (2021). *International Journal of Business Strategy and Automation (pp. 1-19).*
www.irma-international.org/article/pricing-strategies-for-companies-during-the-covid-19-pandemic/287111

Distribution and Selection of Ornamental Fishes' Issues on a Koi Fish Pond Using Krill Algorithm to an Order Picking Model
Erwin Adán Martinez Gomez (2019). *Handbook of Research on Metaheuristics for Order Picking Optimization in Warehouses to Smart Cities (pp. 275-288).*
www.irma-international.org/chapter/distribution-and-selection-of-ornamental-fishes-issues-on-a-koi-fish-pond-using-krill-algorithm-to-an-order-picking-model/227171