Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 279 (2015) 101–111

FUZZY sets and systems

www.elsevier.com/locate/fss

Fuzzy precision and recall measures for audio signals segmentation

Bartosz Ziółko [a,b,*]

[a] *AGH University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, al. Mickiewicza 30, Kraków, Poland*

[b] *Techmo sp. z o.o., Kraków, Poland*

Received 23 December 2013; received in revised form 10 March 2015; accepted 11 March 2015

Available online 14 March 2015

Abstract

The approach presented in this paper applies fuzzy set theory to the evaluation of audio signals segmentation with high resolution and accuracy. The method is based on comparing automatically found boundaries with *ground truth*. Hence, the method is more accurate and able to grasp the evaluation problem in a way more similar to the evaluation conducted by a human being. Traditional methods often fail on grading segmentation algorithms, particularly those of relatively similar qualities. We define a fuzzy membership function that measures the degree to which the segments obtained by an automatic procedure are similar to the results of a correct segmentation. To identify a pair of equivalent segments, we set a fuzzy alignment function that points the pairs of segments obtained by an automatic segmentation with the corresponding segments from a correct segmentation. Speech segmentation is an example where the presented approach was applied.

Keywords: Signal segmentation; Evaluation measures; Speech processing

1. Introduction

The problem of segmentation is an important task in various fields, including natural language [1] and audio signals processing [2–5]. It is a challenging issue, and there are many different approaches. It is not an easy task to evaluate and compare these approaches, due to the ambiguity inherent in classification problems [6]. First, in many cases, there are processes which cannot be clearly compared with the *ground truth*, because it is difficult to perform a totally correct segmentation, even manually or under supervision. A lot of evaluations of segmentation methods are subjective, based either on the direct human decision, or on the supervised evaluation, in which the segmented data are compared against manually segmented or pre-processed data.

* Correspondence to: AGH University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, al. Mickiewicza 30, 30-059, Kraków, Poland.
E-mail address: bziolko@agh.edu.pl.
URLs: http://www.dsp.agh.edu.pl, http://techmo.pl.

http://dx.doi.org/10.1016/j.fss.2015.03.006
0165-0114/

Additionally, the comparison of segmentation results is typically not binary. We can rarely say that the segmentation is definitely correct or definitely wrong. Usually we can say that it is quite accurate or not. This is why fuzzy logic (used in our method) is useful to compare the segmentation results.

A well known classification measure is the 'precision and recall' (P&R) method [7]. It is especially commonly used in information retrieval. Fuzzy logic has already been applied [8] to P&R measure in speech segmentation of phonemes and for some other signal processing tasks [9] like electro-cardiograph analysis [10], image [11] signal reconstruction [12].

However, in case of speech signals segmentation [8], it was only briefly introduced without a proper mathematical description. In this paper this gap is filled.

A lot of evaluation methods are based on computing the number of segments insertions and deletions (I&D). Typically, some tolerance of small errors is used, so a small difference is not considered as an I&D.

2. Speech segmentation

Surprisingly, the evaluation methods for speech segmentation are usually quite simple and do not consider all possible scenarios. Usually, the details of the evaluation method are developed on an ad hoc basis for a given experiment. This means the methods are not widely applicable and they lose accuracy due to the simplifications that are made. Most often the evaluation methods are based on counting the number of insertions, deletions and substitutions (which is so called I&D method) of the automatic segmentations when compared to a hand-checked *ground truth*.

For example, automatic word segmentation [13] was evaluated by counting the number of boundaries for which the time difference between automatic and manual segmentation exceeded thresholds of 35, 70 or 100 ms. The syllable segmentation [14] accuracy was evaluated by counting the number of insertions and deletions within a tolerance of 50 ms before and after a *ground truth* boundary. Some authors do not present any details about such tolerances or do not specify a tolerance but use generally similar methods [15]. The insertion and deletion approach has a few flaws. First, any particular value of tolerance is arbitrary and cannot be chosen on any theoretical basis. It is chosen empirically, quite often with knowledge of the results of a given speech segmentation method. What is more, such methods treat different inaccuracies as simply correct or incorrect detections without considering 'how incorrect' the detection really is. In another example (for syllables) [14], a boundary tolerance is set to 50 ms based on a statistical average of the length of segments. The disadvantage of this approach is that all linguistically motivated speech segments (words, syllables or phonemes) vary considerably in their length. Therefore, a change of 50 ms in a boundary location is not the same for a 100 ms long syllable as for a 300 ms long one. These problems are solved by our evaluation method.

There are other approaches, for example, the R-measure [6]. It increases towards the ideal target-point (100% hit-rate and 0% over-segmentation). However, it is much more inclined to penalise over-segmentation than the traditional measures, e.g. like the F-value [7].

3. Fuzzy segmentation sets

Fuzzy logic and set theory [16,17] are needed for the proper evaluations of many types of segmentation. The detected boundaries may be shifted more or less compared to a manual segmentation. This 'more or less' makes a crucial difference and cannot be mathematically described in Boolean logic. Fuzzy logic provides a method for grading the detected boundary locations in a more sensitive way.

We suggest the extension of the well-known P&R evaluation method, but with elements classified as elements of a fuzzy set. As usual in P&R measurements, one set contains g*round truth* elements and the other is the set of automatically retrieved ones. We calculate an evaluation measure using the number of elements in each set and in their intersection. The comparison of the number of *ground truth* elements and the number of elements in the intersection with the automatically retrieved set gives a value of precision. The comparison of the number of automatically retrieved elements and the intersection with the *ground truth* set gives recall, which is a grade of wrong detections.

Using Boolean logic, information about how many correct segments were found is obtained. By using fuzzy logic, we evaluate not only how many elements were detected, but how accurately they were detected. In this case, fuzzy logic allows us to evaluate not only the number of imperfect detections but also their level of incorrectness. Each retrieved element has a probability factor which represents its level of 'correctness'.