Available online at www.sciencedirect.com

SciVerse ScienceDirect

Computer Speech and Language 28 (2014) 326–341

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

A fuzzy classifier to deal with similarity between labels on automatic prosodic labeling☆

David Escudero-Mancebo*, César González-Ferreras, Carlos Vivaracho-Pascual, Valentín Cardeñoso-Payo

Departamento de Informática, Universidad de Valladolid, 47011 Valladolid, Spain

Received 11 April 2013; received in revised form 11 July 2013; accepted 5 August 2013
Available online 16 August 2013

Abstract

This paper presents an original approach to automatic prosodic labeling. Fuzzy logic techniques are used for representing situations of high uncertainty with respect to the category to be assigned to a given prosodic unit. The *Fuzzy Integer* technique is used to combine the output of different base classifiers. The resulting fuzzy classifier benefits from the different capabilities of the base classifiers for identifying different types of prosodic events. At the same time, the fuzzy classifier identifies the events that are potentially more difficult to be labeled. The classifier has been applied to the identification of ToBI pitch accents. The state of the art on pitch accent multiclass classification reports around 70% accuracy rate. In this paper we describe a fuzzy classifier which assigns more than one label in confusing situations. We show that the pairs of labels that appear in these uncertain situations are consistent with the most confused pairs of labels reported in manual prosodic labeling experiments. Our fuzzy classifier obtains a soft classification rate of 81.8%, which supports the potential of the proposed system for computer assisted prosodic labeling.

Keywords: Automatic prosodic labeling; Fuzzy classification; ToBI

1. Introduction

Prosodic labeling aims to enrich spoken utterances with labels that are representative of the relationship between the prosodic form and function of the constituents of the message. Although the prosodic labeling systems establish clear rules and protocols, the difficulty of the task and the inherent subjectivity of the labelers' judgments cause a high number of inconsistencies. The prosodic labeling systems assume that uncertain situations could appear and reserve special symbols for representing them (like the symbol '?' in ToBI (Beckman et al., 2005)). Nevertheless, apart from these declared uncertain situations, the inter-transcriber tests have shown a relevant number of situations where two different transcribers assign a different label to the same prosodic unit (Escudero et al., 2012). Moreover, the apparent perceptual and acoustic similarity of several pair of labels and their corresponding prosodic units to be labeled is one of the reasons for the uncertain assignment of prosodic labels (Escudero and Estebas, 2012).

☆ This paper has been recommended for acceptance by T. Kawahara.
* Corresponding author. Tel.: +34 666800836.
E-mail addresses: descuder@infor.uva.es (D. Escudero-Mancebo), cesargf@infor.uva.es (C. González-Ferreras), cevp@infor.uva.es (C. Vivaracho-Pascual), valen@infor.uva.es (V. Cardeñoso-Payo).

0885-2308/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.csl.2013.08.001

Fuzzy sets theory (DuBois and Prade, 1980) has been widely used to represent this type of situations where it is difficult to classify a given element into the different possible categories. The prosodic categories in a prosodic labeling system like ToBI are not fuzzy categories because they have a linguistic phonological meaning. Nevertheless, the labels assigned by human or automatic transcribers are uncertain information according to the experimental evidence. In this paper we show how fuzzy sets can be used in situations where assigning a given prosodic unit to a class is difficult and in situations where more than one class could be assigned as there exists a degree of uncertainty.

The difficulties of manual prosodic labeling are projected into automatic prosodic labeling. Prosodic labeling strongly depends on perceptual judgments that must be performed by human transcribers. Nevertheless, automatic prosodic labeling is a need because manual prosodic labeling is a slow and costly process. There are several applications that can benefit from prosodic labeling, but require the processing of huge corpora. Labeling such corpora manually is not affordable. Thus, automatic prosodic labeling systems need to be refined for this type of applications. This paper introduces a new proposal for this: rather than trying to automatically generate one label per word, our proposed algorithm will generate zero or more labels per word, depending on the value of a single α-cut of the fuzzy classifier.

The state of the art on automatic prosodic labeling reports identification rates higher than 90% in binary decisions, referring to determining the presence or absence of accent, boundary or break. Nevertheless, when trying to classify different types of accents, boundaries or breaks, the classification rates dramatically decrease to about 70% (see Section 2.2 for a review of the state of the art). In González-Ferreras et al. (2010), we showed that the reasons for these low accuracy rates are the high similarity among some pair of classes and the imbalanced nature of the prosodic corpora. In this article we show that the use of a fuzzy classifier considerably increases the accuracy rates when soft classification is performed.

Fuzzy classifiers have shown themselves to be useful in various applications, from system control to decision making (Zimmermann, 1999). In this paper we present a fuzzy classifier based on an adaptation of the prosodic event classifier described in González-Ferreras et al. (2012). With the application of fuzzy expert fusion, the output of the base classifiers is interpreted as a confidence value or a membership degree of a given prosodic unit to a prosodic category. We benefit from the fact that each base classifier behaves differently in the task of separating the different pairs of classes. The application of α-cuts allows the assignment of labels with a certain degree of uncertainty. The method allows the performance to be increased by assigning soft labels to some of the prosodic units. The prosodic units where the uncertainty is higher are identified. We show that the pairs of labels involved in uncertain situations are the most confused pairs that appear in perceptual inter-transcriber tests. The solution we propose in this paper is specially interesting in a situation were automatic labelings are being check by human transcribers, since it is easier for a human to select the best label from a list of candidates than replace the automatic label with some other drawn from perception.

The structure of the paper is as follows. First, we review the state of the art on prosodic labeling, automatic prosodic labeling and fuzzy classification. Next, the fuzzy classifier and the experimental procedure are described. Then we show the results and discuss about the application of the prosodic labeling and the uncertainty of the labeling process. Finally, we present some conclusions.

2. State of the art

2.1. Prosodic labeling

The Autosegmental-Metrical (AM) model (Ladd, 1996) has been widely applied in prosodic annotation of spoken corpora. The model was proposed in Pierrehumbert's thesis (Pierrehumbert, 1980). Based on this model, ToBI (Tones and Break Indices) is a framework for transcribing and annotating the prosody of speech. ToBI-based systems have been developed for many languages such as English (Beckman et al., 2005), Spanish (Beckman et al., 2000; Beckman, 2002; Estebas and Prieto, 2009), German (Grice and Benzmüeller, 1995), Japanese (Venditti, 2005), Greek (Arvaniti and Baltazani, 2005), Korean (Beckman and Jun, 2000) and Catalan (Prieto, 2013; Estebas Vilaplana and Prieto, 2010), among others.

The labeling process using ToBI is costly in time and requires the intervention of highly specialized transcribers. However, there may be some inconsistencies in the annotation, as the inter-transcriber consistency rate reveals. For example, in English Pitch Accent labeling the *kappa index* described in the state of the art goes from 0.51 (Yoon et al., 2004) to 0.69 (Syrdal and McGory, 2000). Furthermore, the automatic approaches do not reach a high performance