WiSeBE: Window-Based Sentence Boundary Evaluation

Carlos-Emiliano González-Gallardo[1,2(✉)] and Juan-Manuel Torres-Moreno[1,2]

[1] LIA - Université d'Avignon et des Pays de Vaucluse, 339 chemin des Meinajaries, 84140 Avignon, France
carlos-emiliano.gonzalez-gallardo@alumni.univ-avignon.fr, juan-manuel.torres@univ-avignon.fr
[2] Département de GIGL, École Polytechnique de Montréal, C.P. 6079, succ. Centre-ville, Montréal, Québec H3C 3A7, Canada

Abstract. Sentence Boundary Detection (SBD) has been a major research topic since Automatic Speech Recognition transcripts have been used for further Natural Language Processing tasks like Part of Speech Tagging, Question Answering or Automatic Summarization. But what about evaluation? Do standard evaluation metrics like precision, recall, F-score or classification error; and more important, evaluating an automatic system against a unique reference is enough to conclude how well a SBD system is performing given the final application of the transcript? In this paper we propose Window-based Sentence Boundary Evaluation (WiSeBE), a semi-supervised metric for evaluating Sentence Boundary Detection systems based on multi-reference (dis)agreement. We evaluate and compare the performance of different SBD systems over a set of Youtube transcripts using WiSeBE and standard metrics. This double evaluation gives an understanding of how WiSeBE is a more reliable metric for the SBD task.

Keywords: Sentence Boundary Detection · Evaluation · Transcripts · Human judgment

1 Introduction

The goal of Automatic Speech Recognition (ASR) is to transform spoken data into a written representation, thus enabling natural human-machine interaction [33] with further Natural Language Processing (NLP) tasks. Machine translation, question answering, semantic parsing, POS tagging, sentiment analysis and automatic text summarization; originally developed to work with formal written texts, can be applied over the transcripts made by ASR systems [2,25,31]. However, before applying any of these NLP tasks a segmentation process called Sentence Boundary Detection (SBD) should be performed over ASR transcripts to reach a minimal syntactic information in the text.

To measure the performance of a SBD system, the automatically segmented transcript is evaluated against a single reference normally done by a human. But

I. Batyrshin et al. (Eds.): MICAI 2018, LNAI 11289, pp. 119–131, 2018.
https://doi.org/10.1007/978-3-030-04497-8_10