CrossMark

Training deep neural networks with non-uniform frame-level cost function for automatic speech recognition

Aldonso Becerra[1] ⬤ · J. Ismael de la Rosa[1] ·
Efrén González[1] · A. David Pedroza[1] ·
N. Iracemi Escalante[2]

Abstract The aim of this paper is to exhibit two new variations of the frame-level cost function for training a deep neural network in order to achieve better word error rates in speech recognition. Optimization methods and their minimization functions are underlying aspects to consider when someone is working on neural nets, and hence their improvement is one of the salient objectives of researchers, and this paper deals in part with such a situation. The first proposed framework is based on the concept of extropy, the complementary dual function of an uncertainty measure. The conventional cross-entropy function can be mapped to a non-uniform loss function based on its corresponding extropy, enhancing the frames that have ambiguity in their belonging to specific senones. The second proposal makes a fusion of the presented mapped cross-entropy function and the idea of boosted cross-entropy, which emphasizes those frames with low target posterior probability. The proposed approaches have been performed by using a personalized mid-vocabulary speaker-independent voice corpus. This dataset is employed for recognition of digit strings

✉ Aldonso Becerra
    a7donso@uaz.edu.mx

    J. Ismael de la Rosa
    ismaelrv@ieee.org

    Efrén González
    gonzalez_efren@hotmail.com

    A. David Pedroza
    P.A.D_16@hotmail.com

    N. Iracemi Escalante
    aivinsg_2682@hotmail.com

[1] Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Av. López
     Velarde No. 801, Col. Centro, C.P. 98068, Zacatecas, México

[2] Department of Basic Sciences, Instituto Tecnológico de Pabellón de Arteaga, Carretera a la
     Estación de Rincón KM 1, C.P. 20670, Pabellón de Arteaga, Ags., México