

PBML

The Prague Bulletin of Mathematical Linguistics**NUMBER ??? MARCH 2015 1–21**

QuEst for High Quality Machine Translation**Ergun Biçici^a, Lucia Specia^b**^a ADAPT CNGL Centre for Global Intelligent Content
School of Computing, Dublin City University^b Department of Computer Science
University of Sheffield

Abstract

In this paper we describe the use of QuEst, a framework that aims to obtain predictions on the quality of translations, to improve the performance of machine translation (MT) systems without changing their internal functioning. We apply QuEst to experiments with:

- i. multiple system translation ranking, where translations produced by different MT systems are ranked according to their estimated quality, leading to gains of up to 2.72 BLEU, 3.66 BLEUs, and 2.17 F₁ points;
- ii. n-best list re-ranking, where n-best list translations produced by an MT system are re-ranked based on predicted quality scores to get the best translation ranked top, which lead to improvements on sentence NIST score by 0.41 points;
- iii. n-best list combination, where segments from an n-best list are combined using a lattice-based re-scoring approach that minimize word error, obtaining gains of 0.28 BLEU points; and
- iv. the ITERPE strategy, which attempts to identify translation errors regardless of prediction errors (ITERPE) and build sentence-specific SMT systems (SSSS) on the ITERPE sorted instances identified as having more potential for improvement, achieving gains of up to 1.43 BLEU, 0.54 F₁, 2.9 NIST, 0.64 sentence BLEU, and 4.7 sentence NIST points in English to German over the top 100 ITERPE sorted instances.

1. Introduction

QuEst is a quality estimation framework that offers a wide range of feature extractors that can be used to describe source and translations texts in order to build