LIG System for Word Level QE task at WMT14
Ngoc Quang Luong, Laurent Besacier, Benjamin Lecouteux

To cite this version:
Ngoc Quang Luong, Laurent Besacier, Benjamin Lecouteux. LIG System for Word Level QE task at WMT14. WMT, 2014, Baltimore, United States. hal-02094756

HAL Id: hal-02094756
https://hal.science/hal-02094756
Submitted on 9 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
INTRODUCTION

- Task 2, WMT14: Word-level Confidence Estimation
- New point: MT outputs are collected from multiple translation means (machine and human).
- Our approach: (Figure 1)

FEATURE TYPES (24 IN TOTAL)

- The conventional features (Table 3): work efficiently in our WMT13 submissions and are inherited in this year’s systems.
- The WMT14 features (bold and italic in Table 3): are more specifically suggested by us for this year.

EXPERIMENTAL SETTINGS

AND PRELIMINARY EXPERIMENTS

- Data sets: Description (Figure 2), Statistics (Table 1)

SYSTEM

LIG System for Word Level QE Task at WMT14

Ngoc-Quang Luong
Laurent Besacier
Laboratoire d’Informatique de Grenoble, France
Benjamin Lecouteux

Table 1: Statistics of training, dev and test sets.

WMT14 FEATURES + WMT13 FEATURES (trained with CRF learning algorithm)

FEATURE SELECTION (FS)

- Objectives: filter the most informative features, eliminate the useless ones.
- Sequential Backward Selection
- Best systems: FS_BIN, FS_L1, FS_MULT (Table 2).

FEATURE SELECTION’s help to Integration of several novel features.

Performance Evolution during FS (Figure 4)

- Best subset: Top 18
- Best proposed feature: Occurrence in mult. systems
- More data (WMT13) helps to boost performance
- Future work: research linguistic features, reinforce the segment-level CE, propose the methodology for Sentence CE relied on the word- and segment-level.

Table 2: Pr, Rc and F for labels of all binary, level 1 and multi-class systems, obtained on dev set.

Table 4: Official results of the submitted systems.

SUBMISSIONS AND OFFICIAL RESULTS

- System Average F(%) F(‘OK’)% (%) FS_BIN (primary) 44.4735 74.0961
  FS_L1 31.7814 73.9856
  FS_MULT 20.4953 76.6645
  BL+WMT13(BIN) 44.1074 74.6503
  BL_L1 31.7894 74.0045
  BL_MULT 20.4953 76.6645

CONCLUSIONS AND PERSPECTIVES

- Integration of several novel features.
- Feature Selection’s help to enlighten the valuable features.
- More data (WMT13) helps to boost performance.
- Future work: research linguistic features, reinforce the segment-level CE, propose the methodology for Sentence CE relied on the word- and segment-level.