In phrase-based SMT, weights of the several components are usually estimated via MERT on a development set.

**FACT:** Weights might not generalise well on different domain test sets.

**GOAL:** Readjust the weights to be more appropriate on those sets without the need for specialised data.

---

**Method and results**

This work combines MERT with a perception training to obtain more robust weights.

**IN-DOMAIN TRAINING.** An improvement of more than 2 points of BLEU with respect to the MERT baseline can be obtained.

**OUT-OF-DOMAIN TRAINING.** When using out-of-domain sets in both trainings slight improvements are still observed with the perception.

---

**Methodology**

**Fundamental equation**

\[
T(f) = \hat{e} = \arg\max \log P(e|f) = \arg\max \sum_i \log P(e_i) + \lambda_\alpha \log \lambda + \lambda_\beta \log \lambda,
\]

\[
\lambda_\alpha \log \lambda + \lambda_\beta \log \lambda = \lambda_\alpha \log \lambda + \lambda_\beta \log \lambda = \lambda_\alpha \log \lambda + \lambda_\beta \log \lambda + \lambda_\gamma \log \lambda.
\]

**System development**

After the SMT training, weights are fitted on a development set:

**STAGE 1**

Minimum Error Rate Training.

Fitted weights: \( \hat{X}_0 \)

**STAGE 2**

Perception Training.

Update of each feature weight \( \lambda_i \) for each sentence so that the translation is closer to the best attainable one (see algorithm).

**The algorithm**

**INPUT:** Training data \( \{(f,e)\}_i \), MERT initial weights \( \lambda \), \( N \) epochs, learning rate \( \epsilon \).

**for** each epoch \( n = 1, ... , N \) \n
**for** each example \( f, i = 1, ... , T \) \n
\( \hat{e} \leftarrow \arg\max \left( \log P(e|f) \right) \)

\( \text{tgt: argmax (BLEU(e|f))} \)

if \( \tilde{N}(f, \hat{e}) \neq \tilde{N}(f, \text{tgt}) \) then

\( \lambda_i \leftarrow \lambda_i + \epsilon \Delta h(f, \hat{e}, \text{tgt}, \text{guess}) \)

end if

end for

end for

**GOLD STANDARD (tgt)**

Sentence with the highest (smoothed) BLEU score in the \( \hat{e} \)-best list.

**UPDATE RULE**

Constant update rule only depending on the direction of change:

\[
\Delta h = \arg\max (\tilde{N}(f, \hat{e}) - \tilde{N}(f, \text{guess})).
\]

---

**In-domain TRAINING**

The quality of the translation worsens on development along in-domain training with Trdev while perturbing the weights.

**ON TEST**

Still, the quality improves significantly on out-domain tests:

**Comparison**

**In/Out-domain Training on Out-of-domain N08 TEST**

On an out-of-domain test set, both in-domain (blues) and out-of-domain (reds & greens) perception trainings improve MERT scores. The latter even surpass the fictitious value that MERT would obtain on N08, [N08]...

**Out-of-domain TRAINING**

During the perception training on N06 the quality of the translation is being improved. It gets a stable value over that of MERT on the same data set.

**ON TEST**

The improvement on out-of-domain test sets is even more evident in this case:

**Acknowledgments**

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013 under grant agreement numbers: 279947 and 247702) and from the Spanish Ministry of Science and Innovation (TIN2009-14075-C03).