

ALaSca: an Automated approach for Large-Scale Lexical Substitution

Supplementary Material

Caterina Lacerra , Tommaso Pasini , Rocco Tripodi  and Roberto Navigli 

 Sapienza University of Rome, Department of Computer Science

 University of Copenhagen, Department of Computer Science

{lacerra, tripodi, navigli}@uniroma1.it,
tommaso.pasini@di.ku.dk

Similarity Threshold Tuning

In the following graphs, we report the results for the tuning of similarity thresholds γ and α , grouped by the similarity measure employed. For tuning the threshold γ , we keep fixed the maximum number of allowed substitutes $\alpha = 1$. As one can see in Figure 1, increasing the similarity threshold to accept a candidate substitute does not affect uniformly the performances across configurations, when considering the contextualised similarity. Including the static measure in the combined similarity, instead (Figure 2), leads to more stable results, that increase with γ , with the exceptions of two spikes in the configurations that include LASER.

Contextualised Similarity

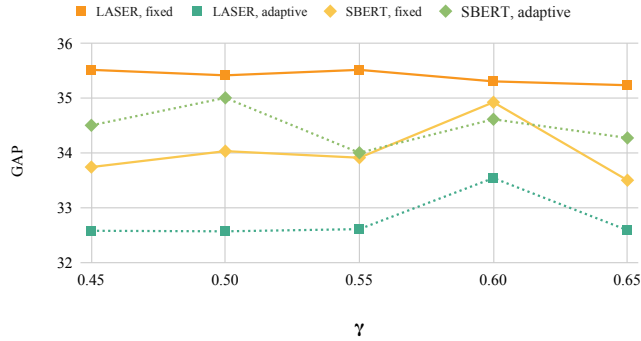


Figure 1: Comparison of changes in GAP scores when using different γ values. Substitutes used in these datasets have been compared with contextualised similarity (crf. Section 3.2 in the paper).

In the second set of experiments, instead, we select the best γ for each configuration and keep it fixed, varying the maximum number of substitutes per instance, i.e., α . In this case, we have a more regular behaviour: both for the configurations that include the contextualised similarity (Figure 3) and for the configurations with the combined similarity (Figure 4), increasing the number of substitutes is beneficial until $\alpha = 3$ in most of the cases; then, continuing adding substitutes does not bring any further improvement but it only brings more noise into the substitutes.

Combined Similarity

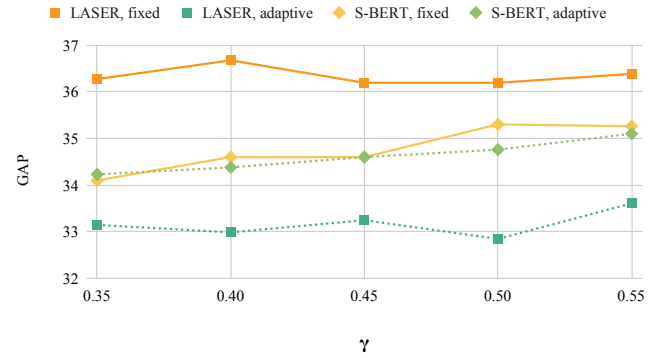


Figure 2: Comparison of changes in GAP scores when using different γ values. Substitutes used in these datasets have been compared with combined similarity (crf. Section 3.2 in the paper).

Contextualised Similarity

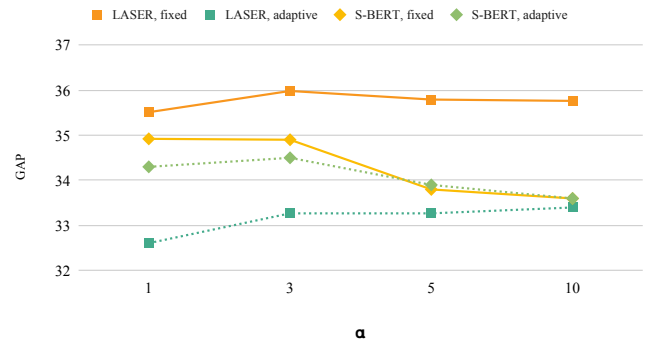


Figure 3: Comparison of changes in GAP scores when using different α values. Substitutes used in these datasets have been compared with contextualised similarity (crf. Section 3.2 in the paper).

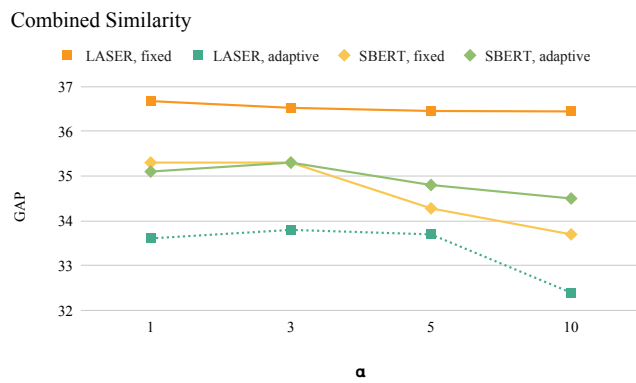


Figure 4: Comparison of changes in GAP scores when using different α values. Substitutes used in these datasets have been compared with combined similarity (crf. Section 3.2 in the paper).