

F1 Scores for Relation Prediction (corrected)

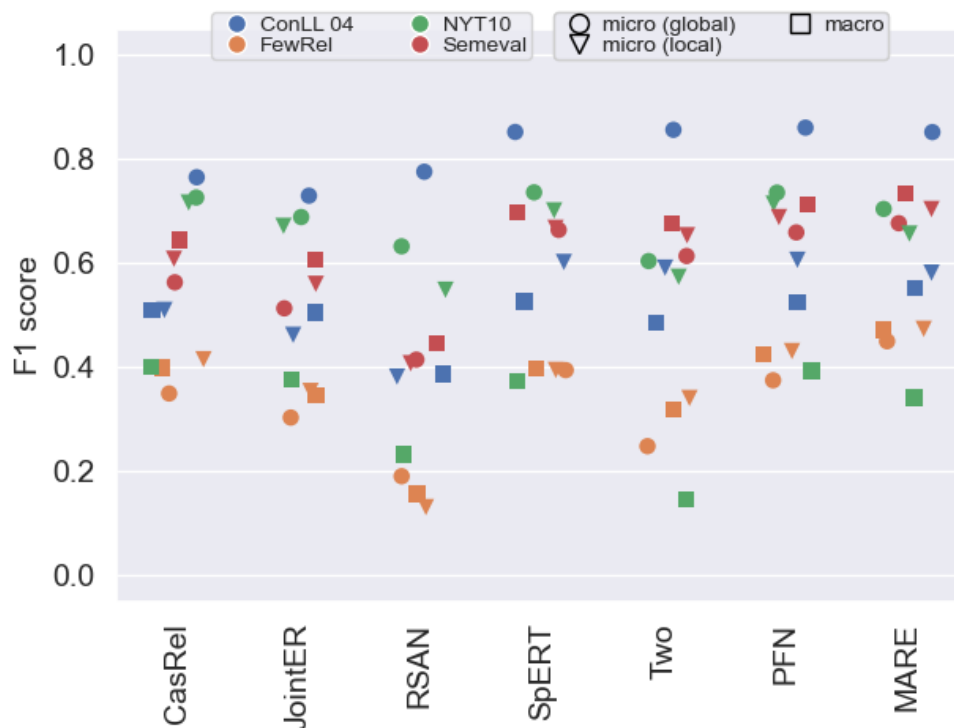


Figure 1 F1 Scores for Relation Prediction, compared to our submission this includes a fix in parsing results, which incorrect lower values

F1 Scores for Entity Prediction Only

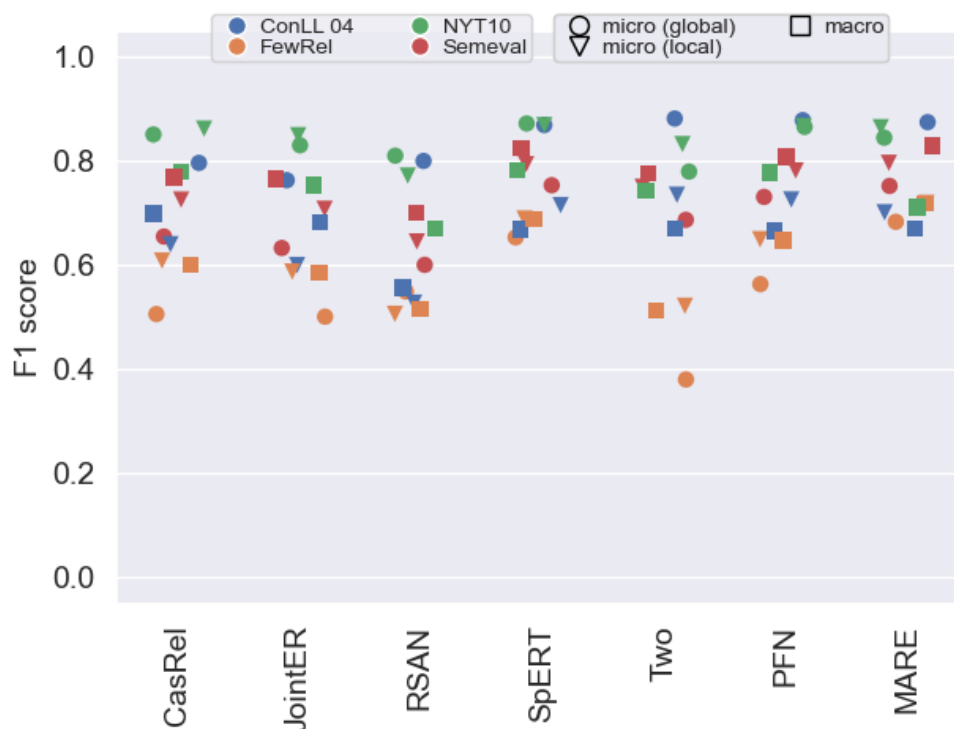


Figure 2 F1 Scores for Entity Prediction by Dataset and Approach

In-Depth Analysis of Model Performances – Input Length

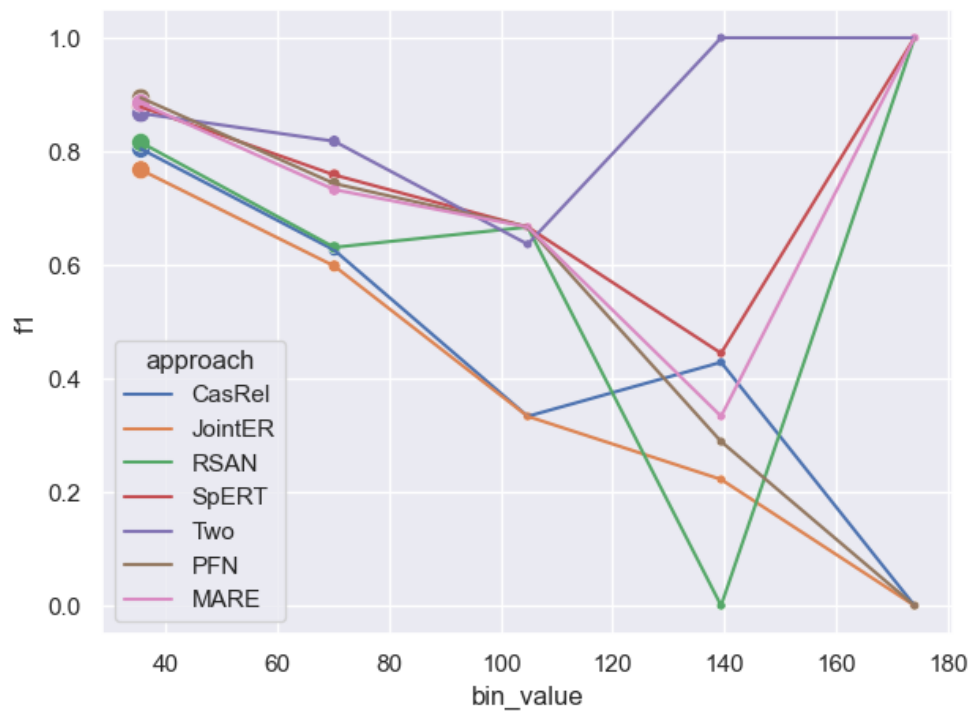


Figure 3 F1 Scores for samples of ConLL 04 binned by their token count. We used 5 bins, size of markers denotes the number of samples present in bin.

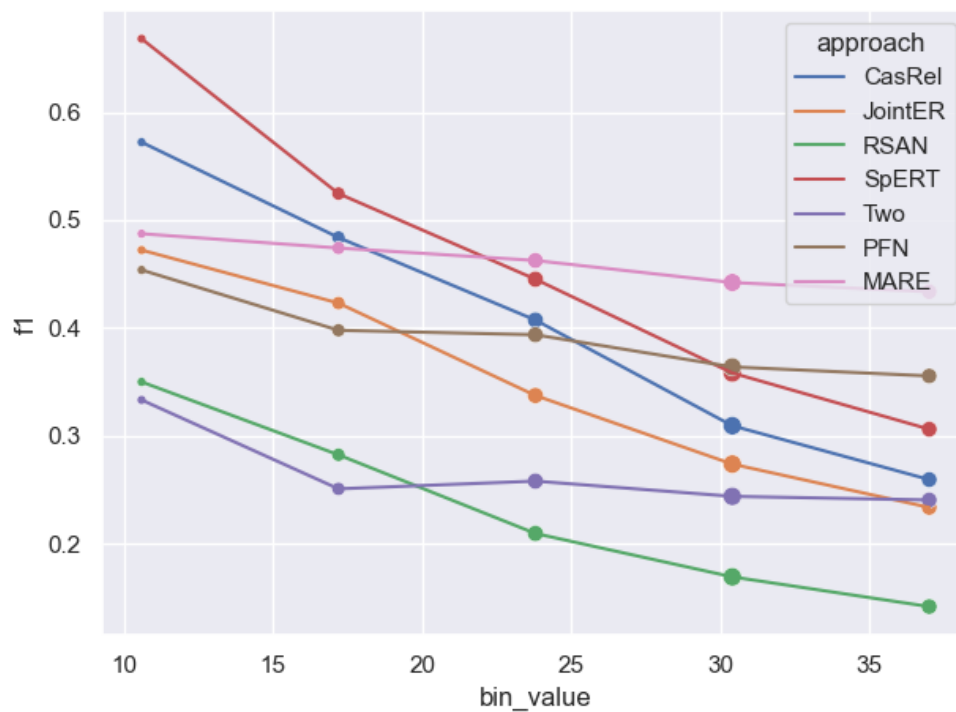


Figure 4 F1 Scores for samples of FewRel binned by their token count. We used 5 bins, size of markers denotes the number of samples present in bin.

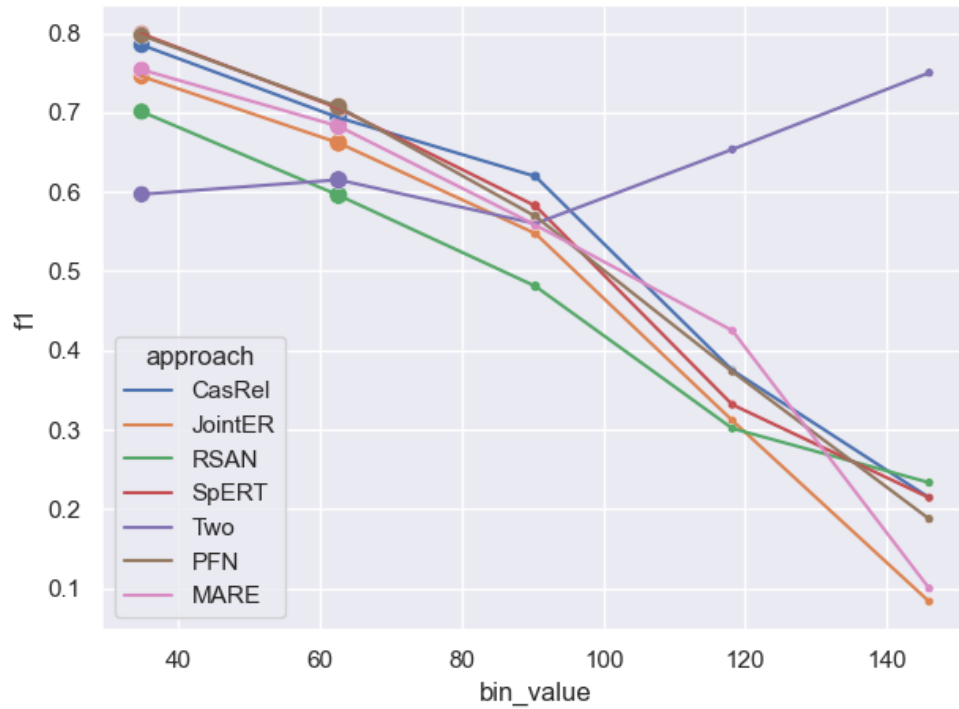


Figure 5 F1 Scores for samples of NYT10 binned by their token count. We used 5 bins, size of markers denotes the number of samples present in bin.

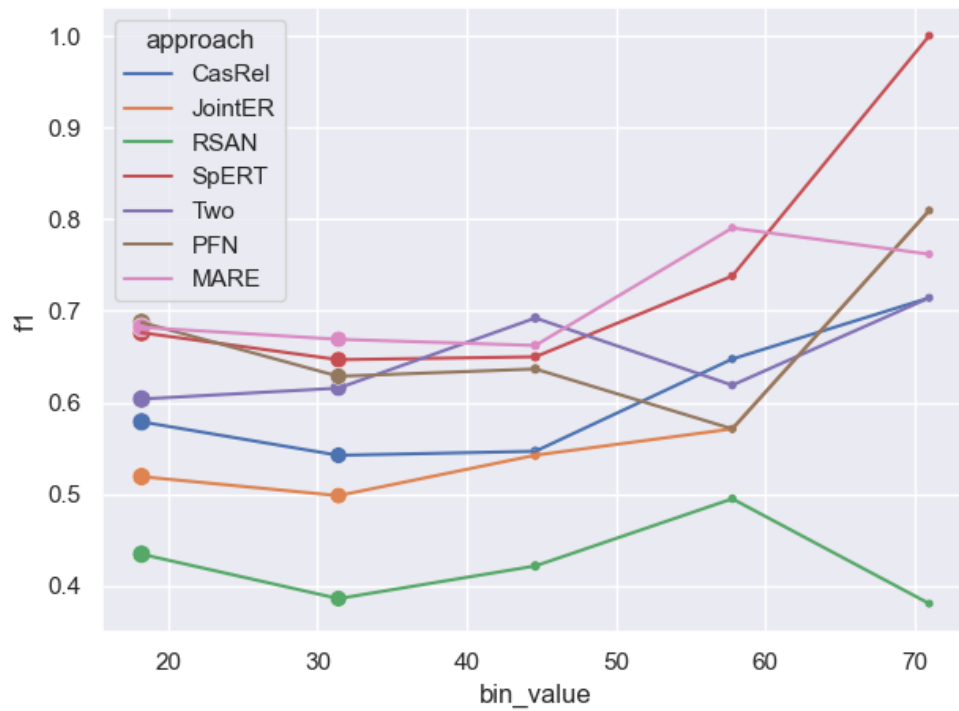


Figure 6 F1 Scores for samples of SemEval binned by their token count. We used 5 bins, size of markers denotes the number of samples present in bin.

In-Depth Analysis of Model Performances – Number of Relations

Note, that only ConLL04 and NYT10 contain samples with different numbers of contained relation mentions.

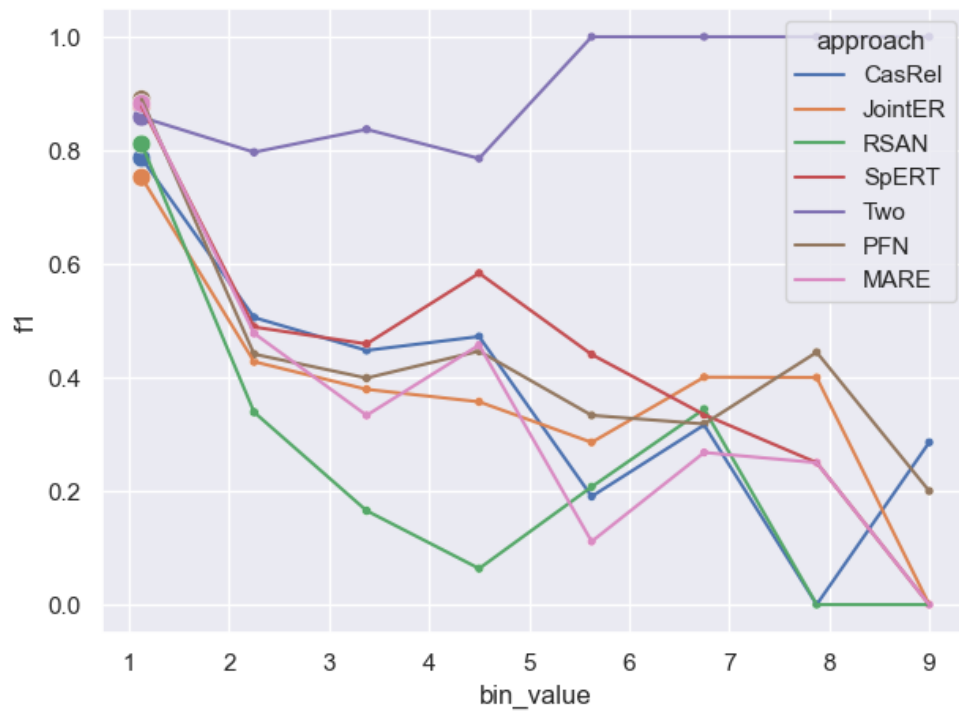


Figure 7 F1 Scores for samples of ConLL04 by number of relations contained in them, size of markers denotes the number of samples present in bin.

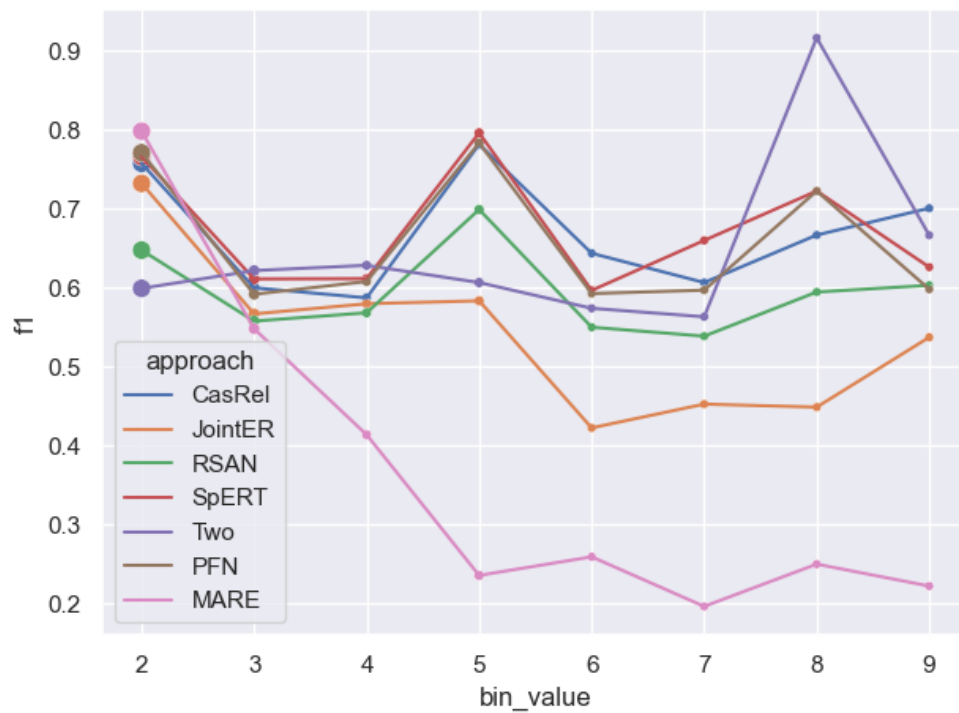


Figure 8 F1 Scores for samples in NYT10 by number of relations contained in them, size of markers denotes the number of samples present in bin.

In-Depth analysis of model performances – Distance between relation

Distance is measured in number of characters between the relations' head and tail entity.

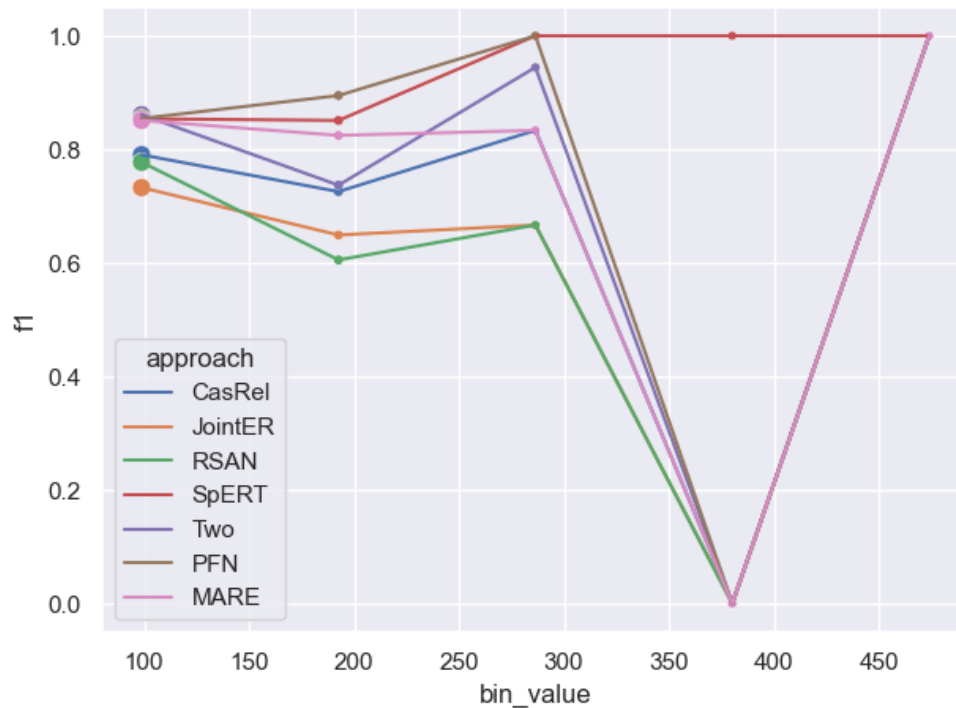


Figure 9 Effect of distance between relation's head and tail entities in ConLL04 on F1 Score, size of markers denotes the number of samples present in bin.

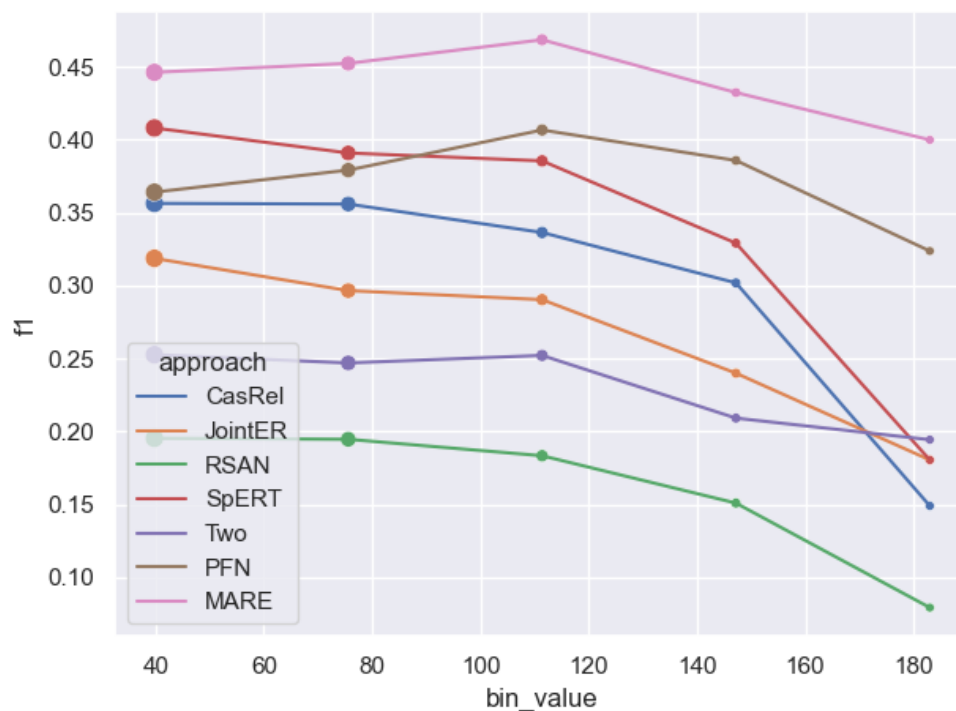


Figure 10 Effect of distance between relation's head and tail entities in FewRel on F1 Score, size of markers denotes the number of samples present in bin.

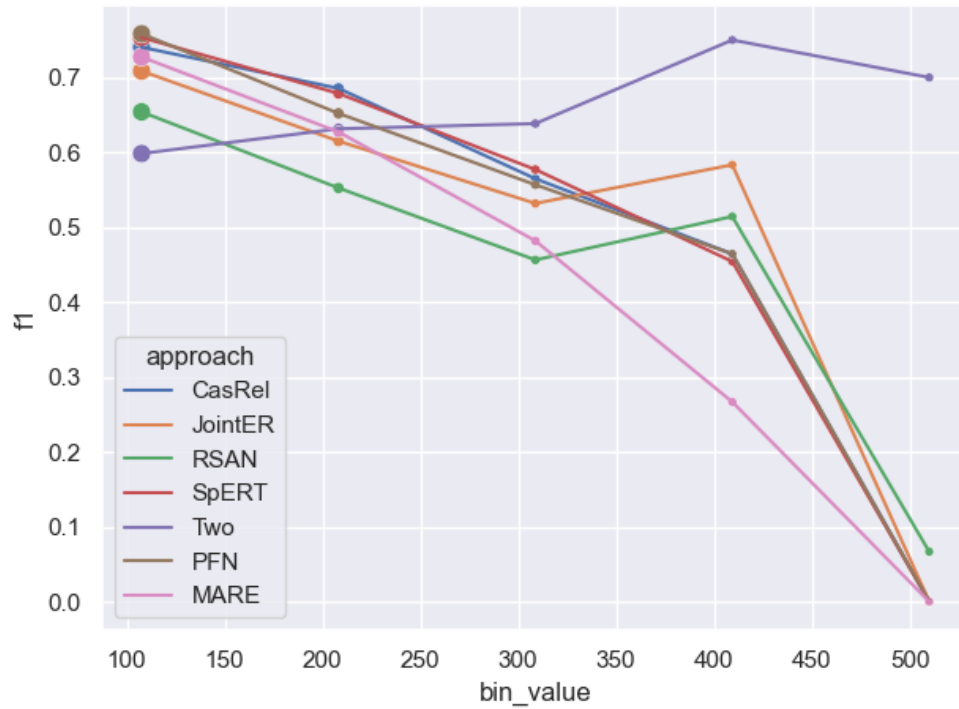


Figure 11 Effect of distance between relation's head and tail entities in NYT10 on F1 Score, size of markers denotes the number of samples present in bin.

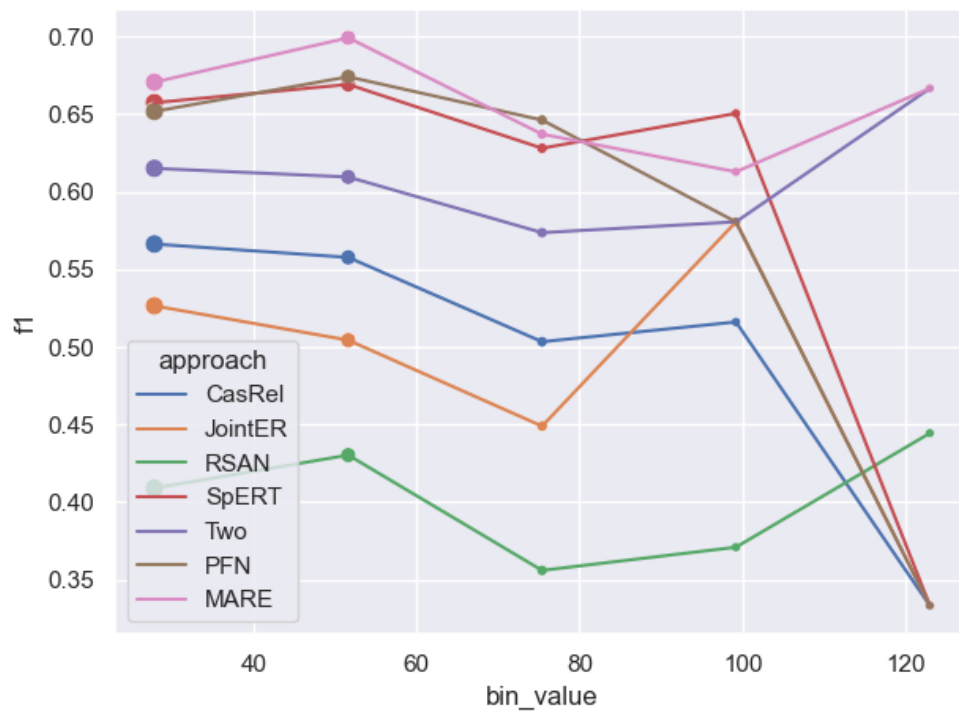


Figure 12 Effect of distance between relation's head and tail entities in SemEval on F1 Score, size of markers denotes the number of samples present in bin.

In-Depth analysis of model performances – Lexical diversity in entity tokens

We measured lexical diversity as the ratio of unique tokens to total tokens used in entity mentions. We scale this ratio to the range [0.0;1.0].

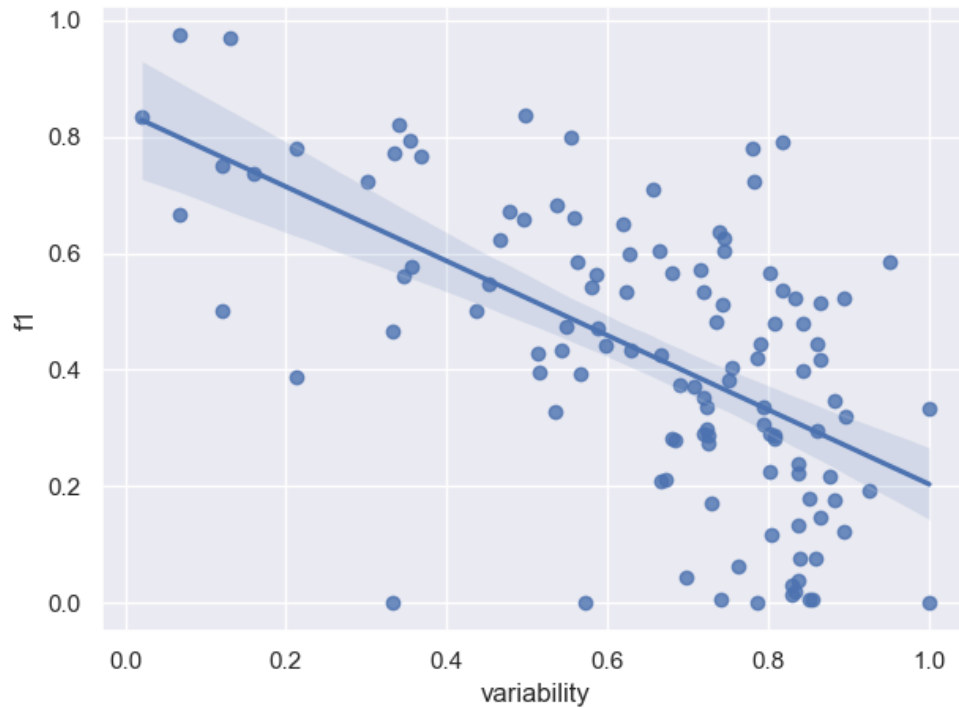


Figure 13 Effect of lexical diversity of entity mentions across all datasets on approach CasRel

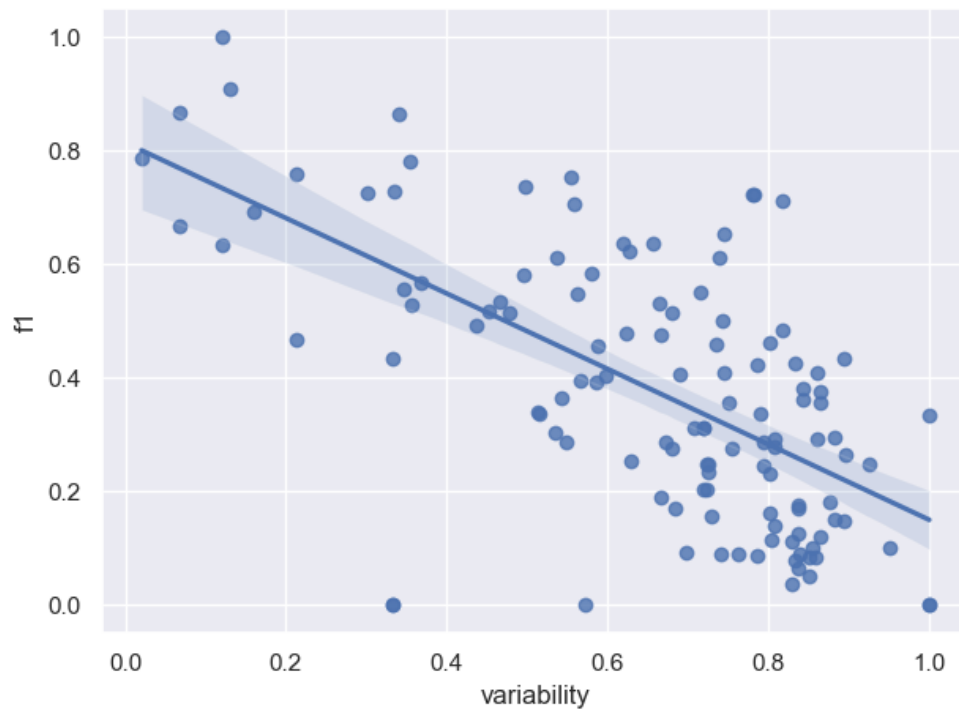


Figure 14 Effect of lexical diversity of entity mentions across all datasets on approach JointER

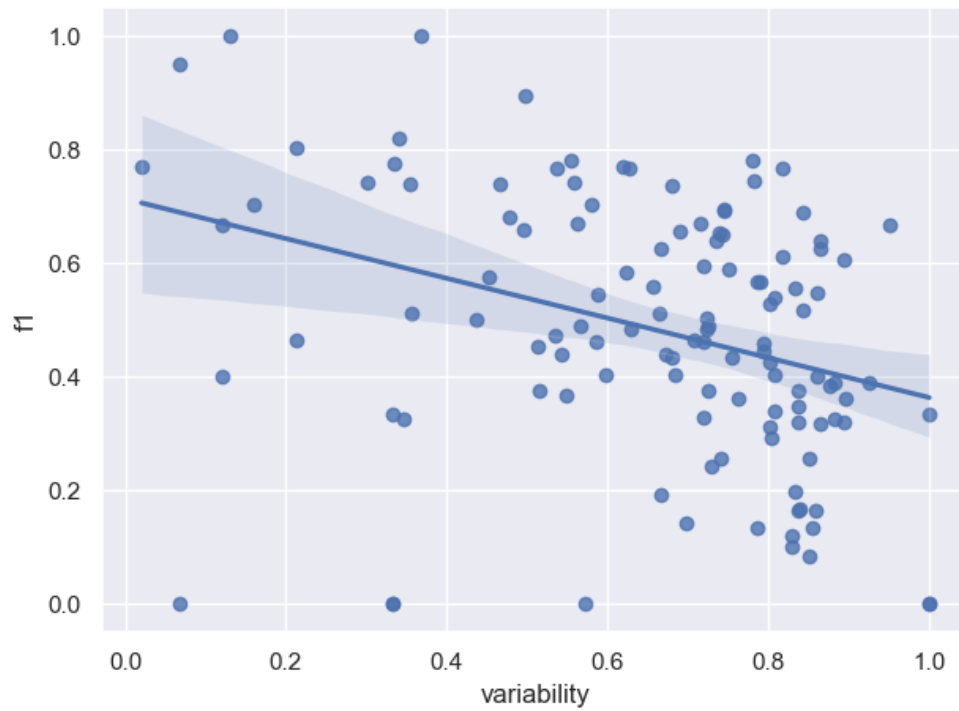


Figure 15 Effect of lexical diversity of entity mentions across all datasets on approach MARE

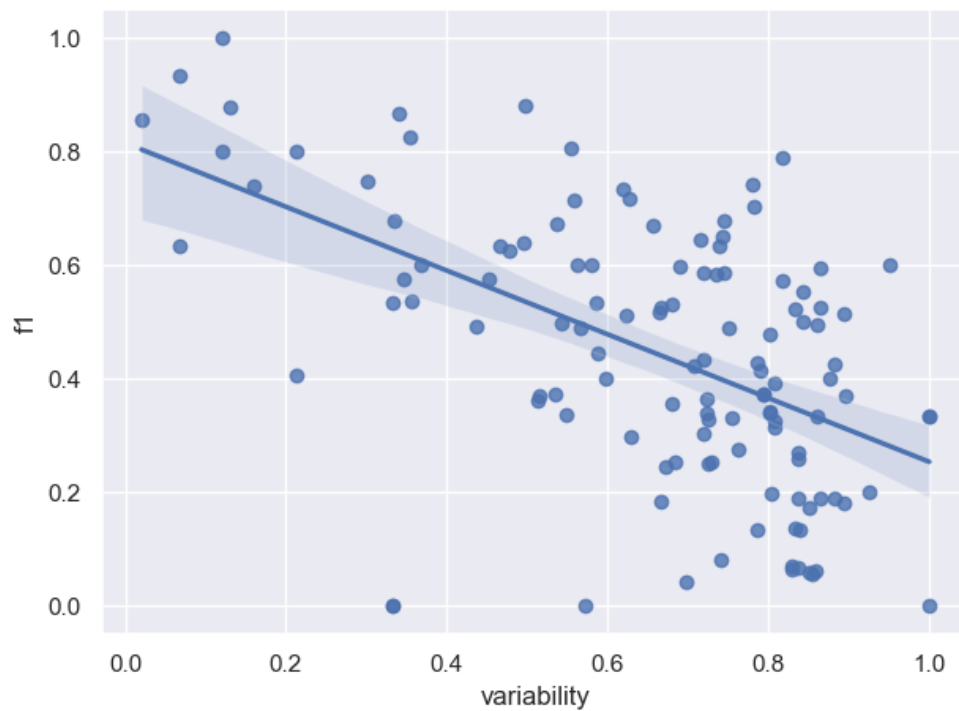


Figure 16 Effect of lexical diversity of entity mentions across all datasets on approach PFN

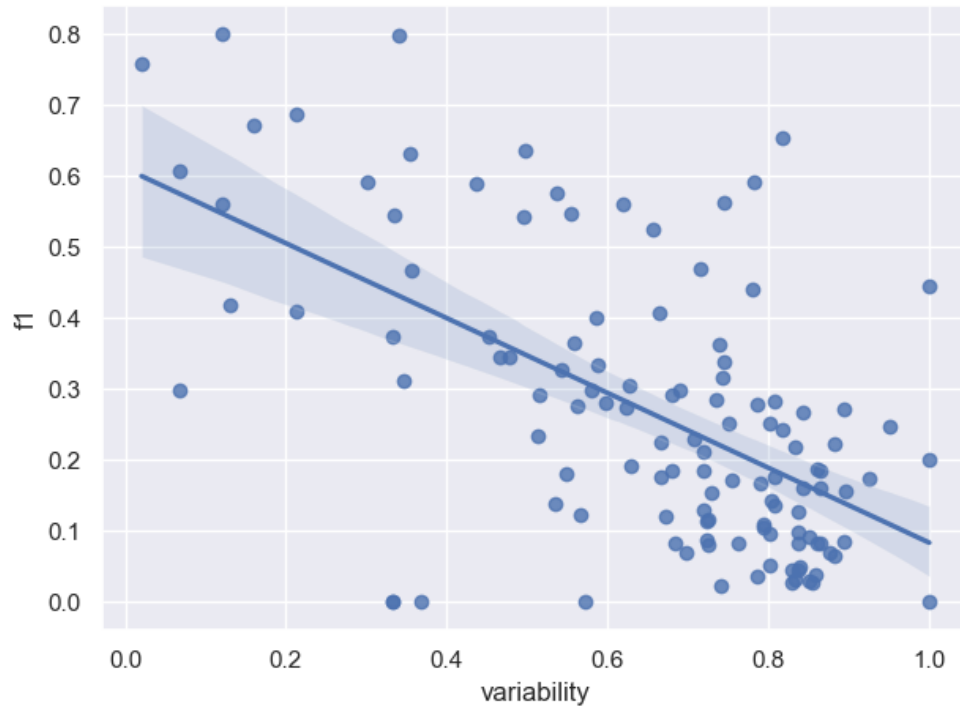


Figure 17 Effect variability in linguistic variability of entity mentions accross all datasets on approach RSAN

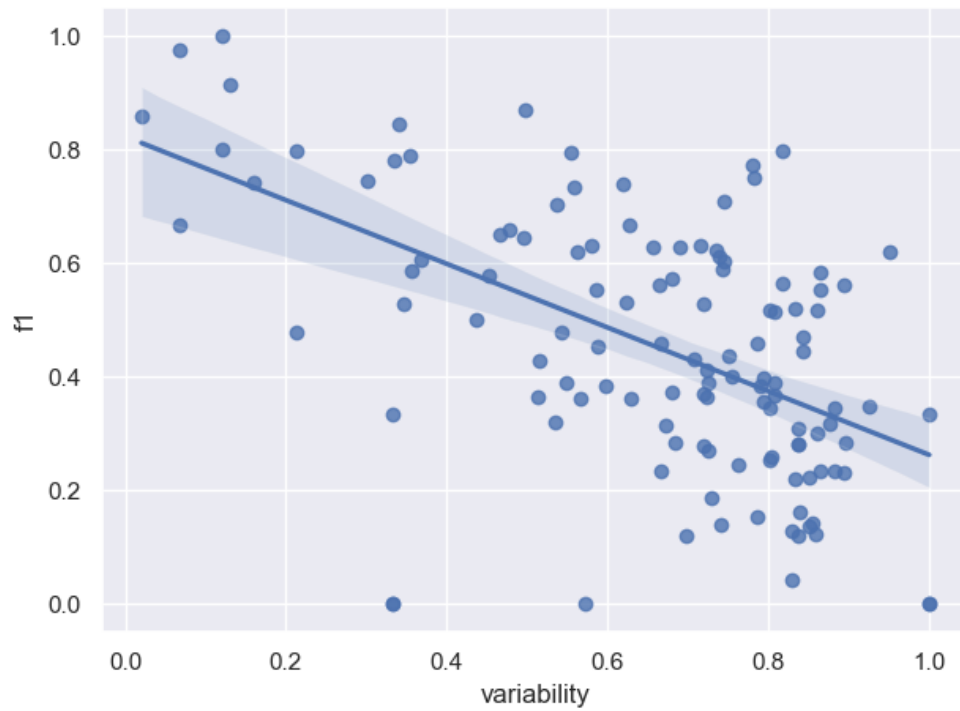


Figure 18 Effect variability in linguistic variability of entity mentions accross all datasets on approach SPERT

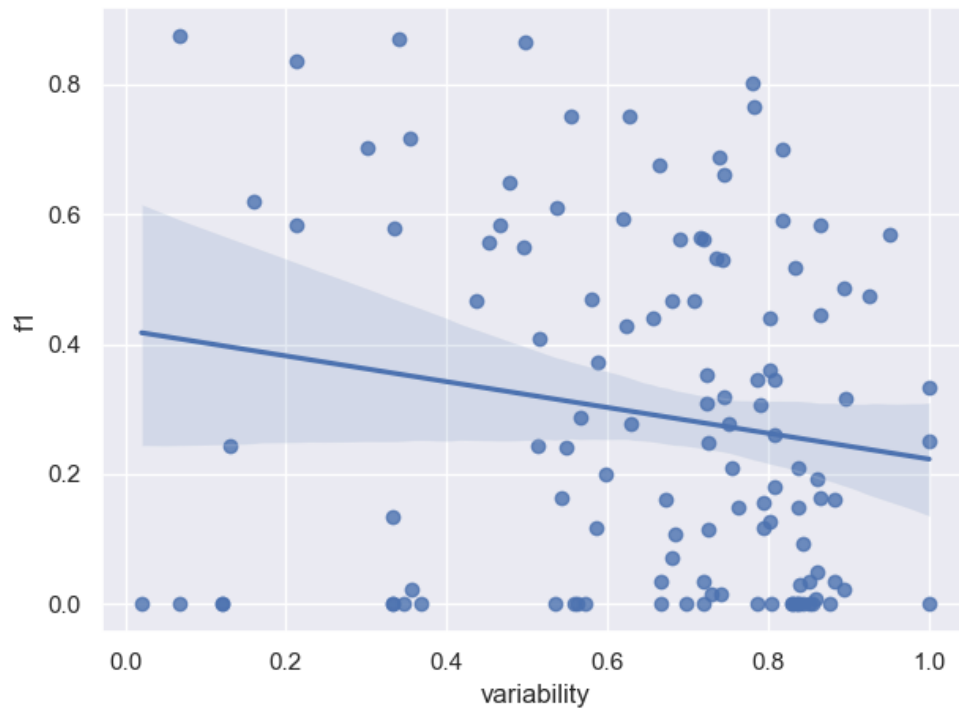


Figure 19 Effect variability in linguistic variability of entity mentions accross all datasets on approach "Two are better than one"