

Word2Vec vs DBnary: Augmenting METEOR using Vector Representations or Lexical Resources?

Additional materials

Christophe Servan^{1,2}, Alexandre Bérard¹, Zied Elloumi^{1,3},
Hervé Blanchon¹ & Laurent Besacier¹

¹LIG – Univ. Grenoble Alpes
Domaine Universitaire
38401 St Martin d’Hères, France
firstname.lastname
@imag.fr

²SYSTRAN
5 Rue Feydeau
75002 Paris, France
firstname.lastname
@systran.fr

³LNE
29 Avenue Roger Hennequin
78190 Trappes, France
firstname.lastname
@lne.fr

Additional results

Tables 1, 3 and 2 presents some additional results to our paper (Servan et al., 2016).

Language pairs	fr-en		de-en		ru-en		Average	
Metric	Threshold	τ	Threshold	τ	Threshold	τ	Threshold	τ
<i>METEOR Baseline</i>	–	0.406±0.015	–	0.334±0.014	–	0.329±0.010	–	0.356±0.013
<i>METEOR DBnary</i>	–	0.408±0.013	–	0.334±0.014	–	0.328±0.010	–	0.357±0.012
<i>METEOR Vector</i>	0.80	0.407±0.014	0.80	0.332±0.014	0.80	0.328±0.010	0.80	0.356±0.013
<i>METEOR Baseline + Vector</i>	0.80	0.407±0.013	0.80	0.343±0.014	0.80	0.332±0.012	0.80	0.361±0.013
<i>METEOR DBnary + Vector</i>	0.80	0.407±0.013	0.80	0.337±0.013	0.80	0.338±0.013	0.80	0.361±0.012
<i>METEOR Vector</i>	0.89	0.411±0.015	0.78	0.333±0.013	0.80	0.328±0.010	0.82	0.357±0.013
<i>METEOR Baseline + Vector</i>	0.73	0.412±0.013	0.80	0.343±0.014	0.88	0.333±0.011	0.80	0.363±0.013
<i>METEOR DBnary + Vector</i>	0.73	0.413±0.013	0.79	0.338±0.013	0.80	0.338±0.010	0.77	0.363±0.012

Language pairs	en-fr		en-de		en-ru		Average	
Metric	Threshold	τ	Threshold	τ	Threshold	τ	Threshold	τ
<i>METEOR Baseline</i>	–	0.280±0.013	–	0.238±0.010	–	0.427±0.013	–	0.315±0.012
<i>METEOR DBnary</i>	–	0.284±0.013	–	0.239±0.010	–	0.435±0.012	–	0.319±0.012
<i>METEOR Vector</i>	0.80	0.290±0.012	0.80	0.241±0.009	0.80	0.436±0.012	0.80	0.322±0.011
<i>METEOR Baseline + Vector</i>	0.80	0.288±0.011	0.80	0.241±0.009	0.80	0.440±0.012	0.80	0.323±0.011
<i>METEOR DBnary + Vector</i>	0.80	0.289±0.011	0.80	0.242±0.011	0.80	0.439±0.011	0.80	0.323±0.011
<i>METEOR Vector</i>	0.72	0.295±0.012	0.79	0.241±0.009	0.72	0.439±0.013	0.74	0.325±0.011
<i>METEOR Baseline + Vector</i>	0.86	0.296±0.012	0.79	0.242±0.010	0.79	0.445±0.011	0.81	0.328±0.011
<i>METEOR DBnary + Vector</i>	0.88	0.294±0.012	0.75	0.245±0.010	0.79	0.443±0.012	0.81	0.327±0.011

Table 1: Correlation score at segment level between several METEOR configurations and human judgments (WMT14 framework). Scores obtained with the *Vector* module are presented firstly with the default threshold (0.80) and secondly with the oracle threshold (under the dashed line).

Models:	<i>monolingual (baseline)</i>	<i>bilingual</i>	<i>pseudo-bilingual with POS</i>	<i>pseudo-bilingual with lemmas</i>	<i>monolingual (lemmas)</i>	<i>bilingual (lemmas)</i>
To English	0.356±0.013	0.354±0.013	0.355±0.013	0.354±0.013	0.357±0.012	0.357±0.012
From English	0.322±0.011	0.322±0.011	0.320±0.011	0.325±0.011	0.324±0.011	0.319±0.010

Table 2: Average correlation score at segment level for *METEOR Vector* with several training configurations of word embeddings with the default threshold (0.80).

Acknowledgements

This work was supported by the KEHATH project funded by the French National Agency for Research (ANR) under the grant number ANR-14-CE24-0016-03.

References

Christophe Servan, Alexandre Bérard, Zied Elloumi, Hervé Blanchon, and Laurent Besacier. 2016. Word2vec vs dbnary: Augmenting meteor using vector representations or lexical resources? In *the proceedings of the 26th International Conference on Computational Linguistics (COLING 2016)*, Osaka, Japan, December.

Language pairs		fr-en		de-en		ru-en		Average	
Metric	Threshold	τ	Threshold	τ	Threshold	τ	Threshold	τ	
<i>monolingual (baseline)</i>	0.80	0.407±0.014	0.80	0.332±0.014	0.80	0.328±0.010	0.80	0.356±0.013	
<i>bilingual</i>	0.80	0.404±0.014	0.80	0.330±0.014	0.80	0.328±0.011	0.80	0.354±0.013	
<i>pseudo-bilingual with POS</i>	0.80	0.409±0.015	0.80	0.333±0.014	0.80	0.322±0.011	0.80	0.355±0.013	
<i>pseudo-bilingual with lemmas</i>	0.80	0.410±0.014	0.80	0.330±0.015	0.80	0.322±0.010	0.80	0.354±0.013	
<i>monolingual (lemmas)</i>	0.80	0.406±0.013	0.80	0.332±0.014	0.80	0.333±0.010	0.80	0.357±0.012	
<i>bilingual (lemmas)</i>	0.80	0.409±0.013	0.80	0.333±0.014	0.80	0.329±0.009	0.80	0.356±0.012	

Language pairs		en-fr		en-de		en-ru		Average	
Metric	Threshold	τ	Threshold	τ	Threshold	τ	Threshold	τ	
<i>monolingual (baseline)</i>	0.80	0.290±0.012	0.80	0.241±0.009	0.80	0.436±0.012	0.80	0.322±0.011	
<i>bilingual</i>	0.80	0.289±0.012	0.80	0.241±0.010	0.80	0.435±0.012	0.80	0.322±0.011	
<i>pseudo-bilingual with POS</i>	0.80	0.284±0.012	0.80	0.241±0.010	0.80	0.434±0.012	0.80	0.320±0.011	
<i>pseudo-bilingual with lemmas</i>	0.80	0.285±0.012	0.80	0.244±0.009	0.80	0.445±0.013	0.80	0.325±0.011	
<i>monolingual (lemmas)</i>	0.80	0.292±0.011	0.80	0.243±0.009	0.80	0.438±0.013	0.80	0.324±0.011	
<i>bilingual (lemmas)</i>	0.80	0.285±0.011	0.80	0.239±0.011	0.80	0.431±0.011	0.80	0.319±0.010	

Table 3: Detail of the correlation score at segment level of the *METEOR Vector* with several training configuration of word embeddings with the default threshold. The last column corresponds to the results show in the Table 2