

1 The evaluation scores for different DBpedia categories of Unseen dataset:

=====

Comparing the evaluation scores between two model: (flat, modify baseline) according to different DBpedia categories, which are: Athlete, Artist, Politician

We mean by "flat" one model from the extended system which contains 3 model: flat, str1, str2

Where we evaluate just this model in term of different DBpedia categories, because the result of other models i.e. str1 & str2 for the whole dataset unseen and unseen not very good comparing by this model

We mean by "Modify baseline" the modified baseline of webnlg challeng

1.1 Table of Evaluation in term of BLEU

Category Name	Model	BLEU
Athlete	Modify Baseline	16.24
	Flat	26.41
Artist	Modify Baseline	6.31
	Flat	14.15
Politician	Modify Baseline	11.12
	Flat	28.65

1.2 Table of Evaluation in term of METEOR

Category Name	Model	METEOR
Athlete	Modify Baseline	0.19
	Flat	0.32
Artist	Modify Baseline	0.13
	Flat	0.27
Politician	Modify Baseline	0.15
	Flat	0.31

1.3 Table of Evaluation in term of TER

Category Name	Model	TER
Athlete	Modify Baseline	0.79
	Flat	0.58
Artist	Modify Baseline	1.03
	Flat	0.78
Politician	Modify Baseline	0.85
	Flat	0.56