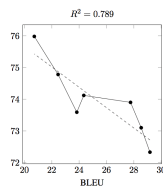
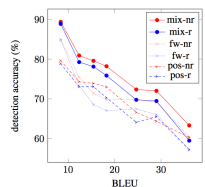


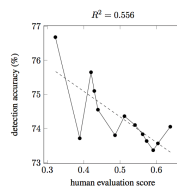
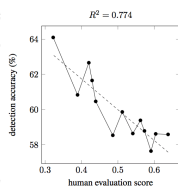
Automatic Detection of Machine Translated Text and Translation Quality Estimation

Roei Aharoni, Moshe Koppel, Yoav Goldberg



	Parallel	Monolingual	BLEU
SMT-1	2000k	2000k	28.54
SMT-2	1000k	1000k	27.76
SMT-3	500k	500k	29.18
SMT-4	100k	100k	23.83
SMT-5	50k	50k	24.34
SMT-6	25k	25k	22.46
SMT-7	10k	10k	20.72

Features	Data	Google	Moses	Systran	ProMT	Linguatrec	Skycode	Trident	R^2
mixed	MT/non-ref	63.34	72.02	72.36	78.2	79.57	80.9	89.36	0.946
mixed	MT/ref	59.51	69.47	69.77	75.86	78.11	79.24	88.85	0.944
func. w.	MT/non-ref	60.43	69.17	69.87	69.78	71.38	75.46	84.97	0.798
func. w.	MT/ref	57.27	66.05	67.48	67.06	68.58	73.37	84.79	0.779
POS	MT/non-ref	60.32	64.39	66.61	73	73.9	74.33	79.6	0.978
POS	MT/ref	57.21	65.55	64.12	70.29	73.06	73.04	78.84	0.948



どんなもの？

分類器の出力を機械翻訳システムの新たな評価指標として提案

入力: 翻訳文の特徴量

出力: 翻訳文が人間由来か機械翻訳由来か

分類器が人間由来と判断するほど良い翻訳文

先行研究と比べてどこがすごい？

参照訳なしで翻訳文を評価できる

考える訳文が多岐にわたる場合でも評価しやすい

技術や手法のキモはどこ？

文の特徴量として, POS tagとfunction wordを使用

分類器はSVM+SMO

どうやって有効だと検証した？

分類精度と既存の評価方法(BLEU)による評価に相関があるか確かめる

➤ 異なるMTシステムや恣意的に翻訳精度を変えたMTシステムでも有効性を確認

分類精度と人間による評価に相関があるか確かめる

➤ 異なるデータセットで有効性を確認

議論はある？

シンタックスベースのMTシステムでは他のMTシステムとは異なる結果になった

POS tagやfunction wordではなくシンタックスベースの特徴であるContext Free Grammarsを特徴量としたところ結果が改善された

次に読むべき論文は？

大学院講義の課題で読んだため特になし