Contents

Chapter 7

Results

After having created a gold standard (see Chapter 6) for evaluating the quality of the alignments, I compared the alignments computed by SimAlign with the alignments computed by a baseline system. I shall now proceed present the results of the experiment.

7.1 Baseline System

As a baseline system, I chose `fast_align` (Dyer, Chahuneau, and Smith 2013). `fast_align` is a re-parameterization of the IBM Model 2. It has become a popular seccessor to Giza++, serves as a baseline system in other works, and is even recommended by WHO? as an alternative for Giza++ for computing the word alignments for Moses SMT. It outperforms Giza++ in many scenarios.

`fast_align` is extremely fast—computing the word alignments for the around 80,000 sentence pairs took around 50 seconds. It is well documented and is extremely easy to compile and to operate. All of this makes `fast_align` the most attractive system to use as a baseline system.

	Method	Dataset Size	Percision	Recall	F_1	AER
Baseline	fast_align	79,109	**0.625**	**0.786**	**0.696**	**0.304**
	”	50k	0.622	0.775	0.69	0.31
	”	25k	0.602	0.751	0.668	0.332
	”	10k	0.58	0.725	0.644	0.355
	”	5k	0.565	0.709	0.629	0.371
	”	600	0.515	0.644	0.572	0.427

Table 7.1: Evaluation metrics for word alignments with the baseline model (fast_align) for different dataset sizes. “Dataset Size” refers to the number of sentence pairs.

	Embedding	Level	Method	Percision	Recall	F_1	AER
SimAlign	mBert	BPE	Argmax	**0.894**	0.622	0.734	0.266
			Itermax	0.832	0.731	0.778	0.222
			Match	0.795	**0.767**	**0.781**	**0.219**
	XLM-R	Word	Argmax	**0.848**	0.399	0.543	0.457
			Itermax	0.767	0.504	0.608	0.391
			Match	0.67	0.647	**0.658**	**0.342**
		BPE	Argmax	0.773	0.488	0.598	0.402
			Itermax	0.671	0.595	0.631	0.369
			Match	0.558	**0.719**	0.628	0.372

Table 7.2: Evaluation metrics for word alignments using SimAlign, with different embeddings and word/sub-word level. Best result per embedding type in bold.

Glossary

Graubünden The Canton of Grisons. 7

Acronyms

AER Average Error Rate. 35

List of Tables

List of Figures

Bibliography

Dyer, Chris, Victor Chahuneau, and Noah A. Smith (June 2013). "A Simple, Fast, and Effective Reparameterization of IBM Model 2". In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia: Association for Computational Linguistics, pp. 644–648. URL: https://aclanthology.org/N13-1073.