Neural Machine Translation by Minimising the Bayes-risk with Respect to Syntactic Translation Lattices

Felix Stahlberg, Adria de Gispert, Eva Hasler, and Bill Byrne

Department of Engineering

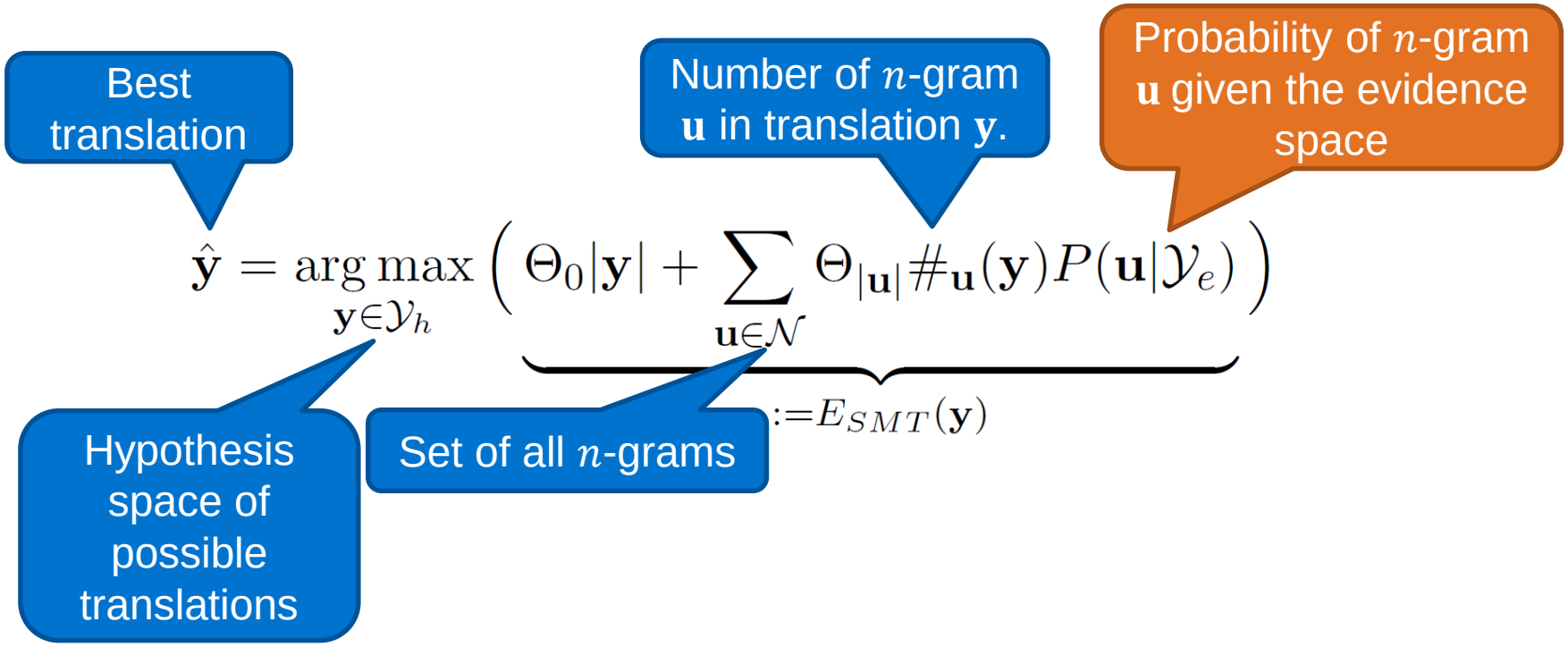
Minimum Bayes-risk decoding in SMT

- Normal decision rule: **maximum a posteriori** (MAP): Select translation with highest probability

vs.

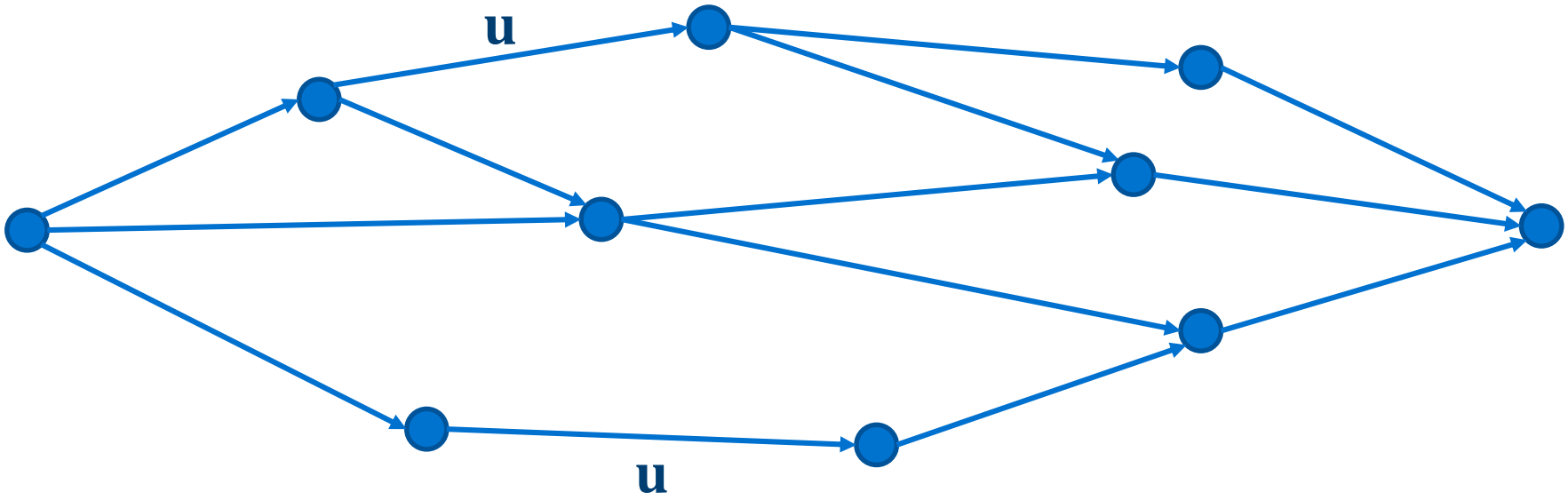
- **Minimum Bayes-risk** (MBR) decision rule: Select translation with lowest expected error in terms of BLEU

MBR decision rule



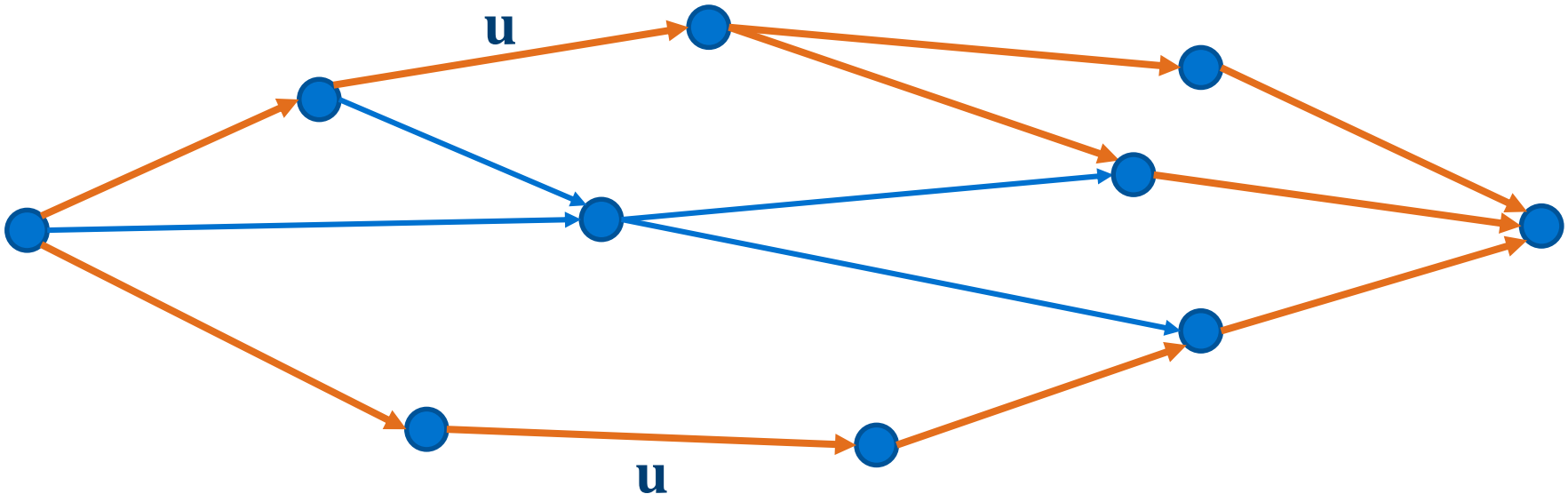
(Kumar and Byrne, 2004; Tromble et al., 2008)

SMT lattices as evidence space



(Tromble et al., 2008; Blackwood et al., 2010)

SMT lattices as evidence space



$P(\mathbf{u}|Y_e)$ = Sum of all orange path probabilities

(Tromble et al., 2008; Blackwood et al., 2010)

Integrating SMT Bayes-risk into the NMT decoder

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} \left(E_{SMT}(\mathbf{y}) + \lambda \log P_{NMT}(\mathbf{y}|\mathbf{x}) \right)$$

Evidence (~Risk) with respect to SMT lattice

Standard NMT translation score

- **Computationally tractable** since risk estimation does not involve NMT.
- **Risk** is computed in a **left-to-right** order.
- The decoder produces n-grams and translations which are **not in the lattice**.
  - ~78% of the translations not in either of the baseline n-best lists.
- The decoder does **not produce UNKs** (UNKs are matched with real words via $E_{SMT}(y)$).

Results on WAT test (Japanese-English)

BLEU scores

	Pure NMT	10k-best Rescoring	This Work (MBR-Based)
SMT Baseline[1]		22.2	
Single NMT (word)	22.5	24.5	**25.2**
6-Ensemble NMT (word)	25.0	25.4	**26.5**
3-Ensemble NMT (BPE)	25.9	25.1	**26.7**

[1] Travatar (Tree-to-string) system (Neubig, 2010)

Results on WMT news-test2015 (English-German)

BLEU scores

	Pure NMT	Lattice Rescoring	This Work (MBR-Based)
SMT Baseline[2]		21.2	
Single NMT (word)	19.6	23.8	**24.6**
5-Ensemble NMT (word)	21.8	24.2	**25.4**
Single NMT (BPE)	21.9	24.0	**24.1**
3-Ensemble NMT (BPE)	23.4	24.3	**24.9**

[2] HiFST (Hiero) system (de Gispert et al., 2010)

Hybrid systems?

Statistical Machine Translation

Neural Machine Translation

System combination
(Ruiz, 2017)

n-best list rescoring
(Neubig et al., 2015)

Discrete SMT-style translation tables in NMT
(Zhang and Zong, 2016; Arthur et al.,2016; He et al., 2016)

Lattice rescoring
(Stahlberg et al., 2016)

NMT features in SMT
(Junczys-Dowmunt et al., 2016)

MBR-based NMT
(this work)

SMT word recommendations for NMT
(Wang et al., 2016)

Symbolic models and neural machine translation

Symbolic Models

Neural Machine Translation

n-best list rescoring
(Neubig et al., 2015)

System combination
(Ruiz, 2017)

Discrete SMT-style translation tables in NMT
(Zhang and Zong, 2016; Arthur et al.,2016; He et al., 2016)

Lattice rescoring
(Stahlberg et al., 2016)

NMT features in SMT
(Junczys-Dowmunt et al., 2016)

MBR-based NMT
(this work)

SMT word recommendations for NMT
(Wang et al., 2016)

References

- Philip Arthur, Graham Neubig, and Satoshi Nakamura. 2016. Incorporating discrete translation lexicons into neural machine translation. In EMNLP, pages 1557–1567, Austin, Texas, USA.
- Graeme Blackwood, Adria de Gispert, and William Byrne. 2010. Efficient path counting transducers for minimum Bayes-risk decoding of statistical machine translation lattices. In ACL, pages 27–32, Uppsala, Sweden.
- Adria de Gispert, Gonzalo Iglesias, Graeme Blackwood, Eduardo R Banga, and William Byrne. 2010. Hierarchical phrase-based translation with weighted finite-state transducers and shallow-n grammars. Computational Linguistics, 36(3):505–533.
- Wei He, Zhongjun He, Hua Wu, and Haifeng Wang. 2016. Improved neural machine translation with SMT features. In AAAI, pages 151–157, Phoenix, Arizona.
- Junczys-Dowmunt, M., Dwojak, T., and Sennrich, R. 2016. The AMU-UEDIN Submission to the WMT16 News Translation Task: Attention-based NMT Models as Feature Functions in Phrase-based SMT. In Proceedings of the First Conference on Machine Translation, Berlin, Germany. Association for Computational Linguistics.
- Shankar Kumar and William Byrne. 2004. Minimum Bayes-risk decoding for statistical machine translation. In HLT-NAACL, pages 169–176, Boston, MA, USA.
- Graham Neubig. 2013. Travatar: A forest-to-string machine translation engine based on tree transducers. In ACL, pages 91–96, Sofia, Bulgaria.
- Graham Neubig, Makoto Morishita, and Satoshi Nakamura. 2015. Neural reranking improves subjective quality of machine translation: NAIST at WAT2015. In WAT, Kyoto, Japan.
- Ruiz, M. 2017 Why Catalan-Spanish Neural Machine Translation? Analysis, comparison and combination with standard Rule and Phrase-based technologies. Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial).
- Felix Stahlberg, Eva Hasler, Aurelien Waite, and Bill Byrne. 2016b. Syntactically guided neural machine translation. In ACL, pages 299–305, Berlin, Germany.
- Roy W. Tromble, Shankar Kumar, Franz Och, and Wolfgang Macherey. 2008. Lattice minimum Bayes-risk decoding for statistical machine translation. In EMNLP, pages 620–629, Honolulu, HI, USA.
- Xing Wang, Zhengdong Lu, Zhaopeng Tu, Hang Li, Deyi Xiong, and Min Zhang. 2016. Neural machine translation advised by statistical machine translation. CoRR, abs/1610.05150.
- Jiajun Zhang and Chengqing Zong. 2016. Bridging neural machine translation and bilingual dictionaries. arXiv preprint arXiv:1610.07272.

Thanks

Code available at http://ucam-smt.github.io/sgnmt/html