

SRInterp: SRI's Scalable Multipurpose SMT Engine

Jing Zheng

Speech Technology and Research Laboratory, SRI International

`zj@speech.sri.com`

Abstract

SRI's statistical machine translation (SMT) engine, SRInterpTM, is designed for both scientific research and real-world applications. SRInterp is equipped with state-of-the-art SMT technology, and has been used in several research and real-time systems for both text and speech translation tasks.

Index Terms: statistical machine translation, real-time system, speech translation

1. Introduction

With the fast-growing demand for global communication and national security, machine translation techniques are receiving more and more research interest. In the past decade, statistical machine translation (SMT) has made very significant progress, and has been widely accepted as the state of the art for machine translation technology, because of its superior performance in many MT benchmarks and applications.

SRI International began its SMT effort under U.S. Defense Advanced Research Projects Agency (DARPA) programs, TRANSTAC (Translation for Tactical Use) and GALE (Global Autonomous Language Exploitation), in 2005, and created SRInterpTM translation technology.

2. SRInterpTM Engine

The SRInterpTM engine is implemented in C++, with careful memory and speed efficiency consideration in algorithm and data structure design. The capability to embrace the latest outcome from SMT research is also a major design goal.

2.1. Translation Model

The SRInterpTM engine supports both phrase-based [1] and hierarchical phrase-based [2] translation models. The former is based on bilingual phrase pairs, and the latter on synchronous context-free grammar rules. Both the phrases and the rules have extensible ASCII formats, which allow researchers to easily plug in new features. The ASCII formats can be bidirectionally converted to a prefix-tree indexed binary format, allowing efficient rule and phrase loading and retrieval in search time. During decoding, translation models can be either stored on disk to minimize memory usage, or loaded into memory to maximize speed.

2.2. Language Model

SRInterpTM uses SRI language modeling (SRILM) [3] libraries to support a large family of n-gram language models, including word-based and class-based, statically and dynamically interpolated models. Extra language model score caching is designed to reduce duplicated computation.

2.3. Search

For phrase-based translation, a multistack dynamic programming algorithm is implemented, using look-ahead score estimation to improve beam search performance. The decoder can optionally store hypotheses back pointers to generate a decoding lattice, from which n-best lists can be extracted with an A* algorithm.

Hierarchical phrase-based translation uses a CKY bottom-up search algorithm supporting target language model intersections. A cube-pruning algorithm [1] is implemented for speed performance. Unique n-best lists can be efficiently extracted from the decoding hypergraph with delayed LM application.

Both search algorithms support high-order language models (as high as 10-gram). N-best lists with kept feature scores can be used for reranking with external knowledge sources.

3. Applications

The SRInterpTM engine has been used in various different systems and applications with different purposes and requirements. In SRI's IraqComm@ Iraqi/English two-way real-time speech-to-speech translation system, translation time of a sentence is typically less than 150 ms in a high-end notebook, which runs three to five translation engines (two are SRInterpTM), two automatic speech recognition (ASR) engines, and two text-to-speech (TTS) engines together in operation mode.

In SRI's GALE evaluation system, translation quality is of the most concern, and the SRInterpTM engine typically uses a large language model with a memory footprint of more than 10 GB.

In the NIST 2008 Open MT evaluation, SRI's SRInterpTM powered systems participated in both the Arabic-to-English (A2E) and the Chinese-to-English (C2E) constrained condition benchmarks. SRI's A2E system scored 42.29% BLEU, ranked 8th out of 18 submissions (6th out of 16 excluding 2 joint entries). SRI's C2E system scored 26.97% BLEU, ranked 6th out of 20 submissions (3rd out of 16 excluding 4 joint entries). A joint submission by MSR, NRC, and SRI ranked at the top (30.89%) among the 20 systems.

4. Conclusions

The versatility and state-of-the-art performance of the SRInterpTM SMT translation engine are demonstrated by several different applications, the recent NIST MT benchmark results.

5. Acknowledgements

Development of the SRInterpTM engine has been supported by DARPA TRANSTAC and GALE programs under contract .

6. References

- [1] Koehn, P., Och, F., and Marcu, D., "Statistical phrase-based Translation," In Proc. NAACL/HLT 2003
- [2] Chiang, D., "A hierarchical phrase-based model for statistical machine translation," in Proc. ACL 2005
- [3] Stolcke, A., "SRILM - an extensible language modeling toolkit," in Proc. ICSLP 2002
- [4] Precoda, K., Zheng, J., Vergyi, D., Franco H., Richey C., Kathol, A., and Kajarekar, A., "IraqComm: A next generation translation system," in Proc. Interspeech 2007