Book Review

Learning Machine Translation

Cyril Goutte[†], Nicola Cancedda[*], Marc Dymetman[*], and George Foster[†] (editors)

([†]Institute for Information Technology, National Research Council Canada; [*]Xerox Research Centre Europe)

Reviewed by
Phil Blunsom
The University of Edinburgh

Attending recent computational linguistics conferences, it is hard to ignore the phenomenal amount of research devoted to statistical machine translation (SMT). Driven by the wide availability of open-source translation systems, corpora, and evaluation tools, a research area that was once the preserve of large research groups has become accessible to those of more modest resources. Although the current state-of-the-art SMT systems have matured into robust commercial systems, capable of providing reasonable quality translations for a variety of domains, they remain limited by naive modeling assumptions and a heavy reliance on heuristics. These limitations have led researchers to ask the question of whether the adoption of techniques from the machine learning literature could allow more complex translations to be modeled effectively. As such, this book, focused on the application of machine learning to SMT, is particularly timely in capturing the current interest of the machine translation community.

Learning Machine Translation is presented in two parts. The first, titled "Enabling Technologies," focuses on research peripheral to machine translation. Topics covered include the acquisition of parallel corpora, cross-language named-entity processing, and language modeling. The second part covers core machine translation system building, presenting a number of approaches applying discriminative machine learning techniques within a SMT decoder.

Much of the content of the book arose from the Machine Learning for Multilingual Access Workshop held at the Neural Information Processing conference in 2006. As SMT is not a frequent topic at that conference, the bridging of research from the mainstream machine learning community with research on MT is particularly promising. A fine example of this cross-over is Chapter 9, "Kernel-Based Machine Translation," in which a novel approach to estimating translation models is presented. However, this promise is not entirely fulfilled, as some contributions either fail to make use of machine learning or are somewhat obscure, unlikely to impact on the mainstream SMT community.

1. Chapter 1: A Statistical Machine Translation Primer

In the first chapter, "A Statistical Machine Translation Primer," the editors seek to both introduce the concept of the book as well as give a brief tutorial on current SMT techniques. In these aims they succeed, describing the elements of current approaches to SMT succinctly. Although those foreign to the field would not come away from reading this chapter able to implement a translation model, pointers to research publications

that contain that level of detail are provided and the authors avoid highlighting obscure research that may mislead.

The introduction also motivates machine translation as an instance of learning with structured outputs, an active area of research in the machine learning community. However, I can't help but feel an opportunity was missed to lay out a clear agenda for research seeking to leverage machine learning techniques in SMT. The issue is not that machine learning can be applied to SMT, but why we would want to do so. Here it is necessary to identify problems with the current approach which could be addressed by a more rigorous statistical treatment: in particular, the lack of structural conditioning and theoretical analysis. Conversely it would seem prudent to highlight the properties of the current approach that have led to its success: the ability to scale to very large corpora and represent phrasal translation units. Trading either of these for more principled learning frameworks is unlikely to yield improvements in performance.

2. Part I (Chapters 2–6): Enabling Technologies

The first section features a collection of five chapters dealing with technologies which are related to, but not core, SMT. Chapter 2, "Mining Patents for Parallel Corpora," by Masao Utiyama and Hitoshi Isahara, describes the application of standard techniques for collecting parallel corpora to a Japanese–English patent data corpus. Lacking any particular novel insights or applications of machine learning, this will mostly be of interest to those seeking data in that particular domain. The problem of constructing dictionaries of named entities and their translation is tackled by Bruno Pouliquen and Ralf Steinberger in Chapter 3, "Automatic Construction of Multilingual Name Dictionaries." This is an interesting problem with relevance for commercial MT systems which must avoid nonsensical literal translations of named entities. However, the treatment takes the form of a system description and fails to make use of any machine learning, thus feeling somewhat out of place in this book.

Things pick up in Chapter 4, "Named Entity Transliteration and Discovery in Multilingual Corpora," by Alexandre Klementiev and Dan Roth, again dealing with named entities but this time making novel use of their temporal occurrence distributions. The authors are able to learn both entity alignments and transliterations with an iterative procedure using the observation that, in temporally aligned parallel corpora, a named entity and its translation will appear co-located in time. This is an interesting technique and should be equally applicable to the alignment of other word types.

Chapter 5, "Combination of Statistical Word Alignments based on Multiple Preprocessing Schemes," by Jakob Elming, Nizar Habash, and Josep M. Crego, tackles the problem of word alignment for morphologically rich languages such as Arabic. To avoid the issue of having to choose a single morphological tokenization, the authors create alignments from a range of tokenizations which are then combined using a binary classifier trained on hand-aligned data. Although of particular interest for those working with Arabic, this chapter fails to go beyond other works on supervised training for word alignment which have consistently shown that it's easy to achieve large gains in alignment accuracy while much more difficult to impact on end-to-end translation performance (Fraser and Marcu 2007).

Part I finishes with a chapter that applies more-advanced machine learning than those before. In "Linguistically Enriched Word-Sequence Kernels for Discriminative Language Modeling," Pierre Mahé and Nicola Cancedda demonstrate the use of string kernels for language modeling, evaluating a number of kernels including one able to integrate a range of factors (surface form, lemma, part-of-speech). This is interesting

work, showing that complex machine learning techniques can be brought to bear on basic NLP tasks, although scaling issues limit the evaluation to small artificial data sets.

3. Part II (Chapters 7–13): Machine Translation

Part II presents a collection of works more directly addressing the title of the book. It is often the case that research seeking to apply machine learning techniques to SMT can neatly be divided into two categories: those that simplify and decompose the translation problem into subtasks that fit existing classification models; and those that maintain the structure of state-of-the-art models and develop new machine learning algorithms specifically for them.

Chapters 7 and 10 fit in the first category. Both decompose the translation problem into subproblems, particularly focusing on lexical choice as classification. In Chapter 7, "Toward Purely Discriminative Training for Tree-Structured Translation Models," Benjamin Wellington, Joseph Turian, and I. Dan Melamed seek to transduce source syntax trees into target strings by learning local classifiers for the nodes in the trees. Although such an approach allows SMT to be viewed as learning local classifiers, the trade-offs made seem to significantly limit the model, something encountered in other works on local tree transduction (Yamada and Knight 2002). In Chapter 10, "Statistical Machine Translation through Global Lexical Selection," Srinivas Bangalore, Stephan Kanthak, and Patrick Haffner take a bag-of-words approach, ignoring ordering information and learning classifiers that predict the presence of target lexical items given an entire source sentence. This chapter takes quite a novel finite–state transducer approach to SMT; however, again the simplifying modeling assumptions seem limiting.

Perhaps the most novel and interesting chapter of this book is Chapter 9, "Kernel-Based Machine Translation," by Zhuoran Wang and John Shawe-Taylor. This work directly addresses the aim of the book: applying powerful state-of-the-art machine learning approaches to machine translation. The authors describe a class of bilingual string kernels capable of modeling phrase-based SMT without constraining phrase extraction with word alignments, instead modeling unrestricted phrase co-occurrence. The learning objective chosen is to minimize the squared loss of the n-gram overlap of candidate translations given the reference, a close fit to the evaluation metric BLEU. The inevitable scaling problems are tackled with a novel information retrieval approach. For each test sentence the algorithm sub-selects training samples based on lexical overlap and decodes using a regression model based on this subset. The results achieved are surprisingly competitive with a standard phrase-based model, an encouraging outcome given that no explicit language model is present in the kernel-based decoder.

Additional chapters (8: "Reranking for Large-Scale Statistical Machine Translation" by Kenji Yamada and Ion Muslea; 11: "Discriminative Phrase Selection for SMT" by Jesús Giménez and Lluís Màrquez; 12: "Semisupervised Learning for Machine Translation" by Nicola Ueffing, Gholamreza Haffari, and Anoop Sarkar) cover relatively well-trodden ground, taking standard SMT models and applying common machine learning algorithms to a sub-part of the system (re-ranking, discriminative phrase selection, and semi-supervised learning, respectively). These chapters provide solid descriptions of applying these techniques and the performance gains that can be achieved, a useful contribution for anyone seeking to augment their existing decoder. However, a caveat here is the evaluation in Chapter 11. Although the authors must be commended on their thoroughness, the vast number of metrics used (one table includes 37!) provides more confusion than clarity when seeking to understand the performance of their system.

In the final chapter, "Learning to Combine Machine Translation Systems," Evgeny Matusov, Gregor Leusch, and Hermann Ney introduce a novel approach to learning system combination models based on confusion networks. This chapter provides a nice treatment of this topic with an evaluation demonstrating the consistent performance gains that can be achieved; it will be of particular interest for those involved in multi-site evaluation campaigns.

4. Summary

In an age in which most research publications can be readily accessed for free via the Web, a collected-works publication such as this stands on its ability to bring together articles which compactly summarize and define a direction of research. In this respect, this book falls short of being a must-buy for the SMT researcher, as many of the works tend towards the esoteric, making it hard for someone seeking familiarity with the field to separate core contributions from those unlikely to represent its future. However, the high degree of novelty and range in the collected articles, with many authors proposing new structures for translation models, still make this a worthwhile read with great potential to inspire future research.

References

Fraser, Alexander and Daniel Marcu. 2007. Measuring word alignment quality for statistical machine translation. *Computational Linguistics*, 33(3): 293–303.

Yamada, Kenji and Kevin Knight. 2002. A decoder for syntax-based statistical MT. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 303–310, Philadelphia, PA.

Phil Blunsom is a research fellow in the School of Informatics at the University of Edinburgh. He conducts research focusing on the application of machine learning to complex structured problems in language processing, such as machine translation, language modeling, parsing, and grammar induction. Blunsom's address is School of Informatics, The University of Edinburgh, 10 Crichton Street, Edinburgh, EH8 9AB, United Kingdom; e-mail: pblunsom@inf.ed.ac.uk; URL: `http://homepages.inf.ed.ac.uk/pblunsom`.