Mach Translat (2009) 23:269–271
DOI 10.1007/s10590-010-9069-2

BOOK REVIEW

Review of Cyril Goutte, Nicola Cancedda, Marc Dymetman, and George Foster (eds): Learning machine translation

MIT Press, 2009, xii+316 pp

Philipp Koehn

Received: 6 January 2010 / Accepted: 16 January 2010 / Published online: 7 February 2010

As the field of statistical machine translation matures, there is not only a need to develop better linguistic models but also better parameter estimation methods. Not only must we develop models that match more closely the linguistic processes to explain the transformation of a text in one language into a meaning-equivalent text in another. There is also a need to refine the core methodology of obtaining parameter values for such models.

The book *Learning Machine Translation* is a timely collection of articles that focus on this second task. The book is part of the *Neural Information Processing Series (NIPS)* of *The MIT Press*, and it is testament to the fact that researchers whose main focus is machine learning have embraced the challenges of machine translation—a development that will be beneficial to both researchers in machine learning and machine translation.

The book opens with a good general overview of current methods in statistical machine translation which touches on all major aspects. It gives sufficient background to readers who are not familiar with the field and prepares them for the more detailed following chapters. The following 12 articles are grouped into two parts: one on enabling technologies and one on machine learning approaches to machine translation. Half of these papers were presented in an earlier form at a 2006 NIPS workshop, while the others stem from newer research.

The first part, which covers parallel corpus acquisition, creation of multi-lingual name dictionaries, named entity transliteration, use of morphological analysis to improve word alignment, and discriminative language modeling is a diverse set of papers. Some of them describe rather narrow research projects (Chaps. 3, 4, and 5),

P. Koehn (✉)
School of Informatics, University of Edinburgh, Edinburgh, UK
e-mail: pkoehn@inf.ed.ac.uk

and overall they do not seem to fit well into the collection. They are neither representative of the field as a whole, nor especially outstanding articles of general interest.

The focus and strength of the book are the seven articles that give a very broad overview of the various angles of applying machine learning to the machine translation problem.

Maybe the machine learning idea with the most history is re-ranking the *n* best translations produced by a traditional statistical machine translation system. Kenji Yamada and Ion Muslea present such work based on a perceptron algorithm, showing minor gains.

Also processing the output are Evgeny Matusov, Gregor Leusch, and Hermann Ney, who report on the combination of multiple machine translation systems using confusion networks. This work represents a recent and very popular trend in the field, which has sparked much interest and even controversy due to the large gains that have been observed (some are not convinced that the observed gains in automatic scores are true improvements in translation quality, and some are worried that just collecting diverse systems for combinations replaces careful examinations of specific models).

A third article that leaves the core machine translation system unchanged, by Nicola Ueffing, Gholamreza Haffari, and Anoop Sarkar, concerns itself with semi-supervised learning for machine translation.[1] They show that useful additional training data can be generated by translating in-domain source language texts and filtering the resulting sentence pairs to keep only the most reliable ones.

An intermediate step towards restructuring machine translation approaches around more advanced machine learning methods is to use them only for components of the traditional models. Along these lines, Jesús Giménez and Lluís Màrquez present work on a discriminative phrase selection model that incorporates context features. While in traditional statistical machine translation systems, scores for translating an input phrase (any sequence of words, not necessarily a syntactic constituent) into an output phrase are based on frequencies in the training data, in a discriminative approach, these scores are optimized to improve translation performance.

Finally, there are three articles that a leap ahead and present very different approaches to machine translation. None of them deliver at this point performance close to the state-of-the-art methods, but open up new possibilities.

Zhuoran Wang and John Shawe-Taylor use kernel-based methods in a phrase-based model to estimate phrase translation scores. However, the use of a non-linear kernel causes problems in decoding (partial scores cannot be simply multiplied together) and results are only reported on a small reduced-domain corpus.

Srinivas Bangalore, Stephan Kanthak, and Patrick Haffner propose an even more radical departure from traditional approaches: Their global lexical selection model predicts the output words and phrases from the entire source sentences, yielding a bag of words from which a sentence has to be reconstructed. Their experiments show that good words are proposed (as measured by precision and recall), but the overall BLEU scores are drastically lower for any but the shortest sentences.

[1] Editor's note: much of the material in this chapter had already been published in 2007 by the same authors in this journal, volume 22, issue 2, pages 77–94.

Benjamin Wellington, Joseph Turian, and I. Dan Melamed aim at a purely discriminative training method for tree-based translation models. They also start with a bag-of-word translation model and limit tree transformation to children reordering and leaf substitutions.

The use of machine learning methods to improve machine translation is a very active and fast-moving subfield. It is therefore not surprising that the book is missing some more recent work that has become influential, such as David Chiang's use of the MIRA algorithm to support a large number of features in the tuning stage (Chiang et al. 2009)—"11,001" instead of the typical 10–20—or Phil Blunsom's work on Bayesian methods for synchronous grammar induction (Blunsom et al. 2009).

In conclusion, the book gives a good starting point to anybody interested in improving machine learning methods for problems such as machine translation, or improving machine translation with machine learning. However, since most of the content of the book are detailed articles on very specific methods, it does not serve as a general introduction to readers who have only very limited background in machine learning or machine translation.

References

Blunsom P, Cohn T, Dyer C, Osborne M (2009) A Gibbs sampler for phrasal synchronous grammar induction. In: Proceedings of the joint conference of the 47th annual meeting f the ACL and the 4th international joint conference on natural language processing of the AFNLP, Suntec. Association for Computational Linguistics, Singapore, pp 782–790

Chiang D, Knight K, Wang W (2009) 11,001 new features for statistical machine translation. In: Proceedings of human language technologies: the 2009 annual conference of the North American chapter of the association for computational linguistics. Association for Computational Linguistics, Boulder, Colorado, pp 218–226