DESIDOC Journal of Library & Information Technology, Vol. 30, No. 4, July 2010, pp. 25-32

Statistical Machine Translation

Mukesh, G.S. Vatsa, Nikita Joshi, and Sumit Goswami**

DESIDOC, DRDO, Metcalfe House, Delhi-110 054
E-mail: mukesh@desidoc.drdo.in; sushil1951@rediffmail.com

**E-mail: nikitajoshi502@gmail.com*

***DRDO HQ, DRDO Bhavan, Delhi-110 105*
E-mail:sumit_13@yahoo.com

ABSTRACT

Statistical Machine Translation (SMT) systems are based on bilingual sentence aligned data. The quality of translation depends on the data provided for translation learning. A huge parallel corpus is required for performing the statistical machine translation. The aim of this paper is to explore SMT using the Moses toolkit for creating a German-English translator. To perform the German to English translation, a parallel corpus of this language pair has been provided. Larger the size of the data provided for the training of the Moses decoder, more accurate is the translated output.

Keywords: Statistical machine translation, machine learning, natural language processing, bilingual corpus

1. INTRODUCTION

Machine translation (MT) is unable to achieve high accuracy because natural languages are highly complex. When taken at the word level, one has to deal with the problem of synonymies, i.e., one word having multiple meanings and different possible translations. At the syntactic level, sentences might have different renderings, and relationship between linguistics entities is often vague. Accurate translation, which is sometimes not possible even for human translator, also encompasses world knowledge or commonsense into account. The number of relevant dependencies is very large and these dependences are complex for a machine translation system to be taken into account. The main focus of MT research until the end of 1980s was mainly using linguistic rules of various kinds such as lexical rules, morphological rules, rules for syntactic analysis, rules for lexical transfer, rules for syntactic generation, rules for morphology, etc. The rule-based approach was most commonly used in the dominant transfer system and was also the basis of various Interlingual systems.

Some improvement was achieved by the development of improved computer hardware, programming languages, and above all the progress in syntactic analysis based on research in formal grammars by Chomsky[1-3] and others. This lead to relegation of this approach by the emergence of new 'corpus-based' methods and strategies. An important development has been the revival of the statistic-based approaches, which were virtually used as the sole means of analysis and generation and where no linguistic rules were applied. SMT is by far the most dominant paradigm of machine translation. The idea behind statistical machine translation is derived from the fact that the document is translated on probability, i.e., a phrase or sentence in language 'B' is the translation of a sentence or phrase in language 'A'.

Improvements in this field were also the result of broad-based research, building upon a wide range of well-tested computational and linguistic methods and techniques. With research in the field of artificial intelligence and natural language processing in the early 1990s, there was an increase in the quantity of bilingual data available on Internet which could be used to train translation systems. Much of the early statistical work used the English-French translations of the Canadian parliament. Similar parallel texts of European Parliament (Europarl)[4-5] transcripts provided training data for other language pairs. Most research was focused on the

Received on 11 January 2010

corpora-based methods, and on combining these with traditional rule-based methods. Empirical translation research has made strides from IBM's Candide system[6] to the latest Google translation services. The research program of statistical machine translation has focused on building statistical models of existing translation data, and on finding optimal translations of new-source-text using these models. A source text with free or erroneous multiple human translations is more helpful than a single translation. Multiple translations can help us in giving more information about the variation of translations, and how the same text has been translated in exactly the same context.

But the availability of parallel texts in the language pair or domain remains questionable. To overcome this shortfall, researchers have turned to comparable corpora such as technical articles on similar topics. Words with similar distributions in comparable corpora have more probability of being translated. The majority of these technical literature is not translated. Subject ontologies and overlapping technical terms provide an aid to identify comparable articles in different languages. More terms can then be bootstrapped into the translation dictionaries and the process is repeated.

The following steps are taken in building data-driven MT systems:

(i) Alignment finds articles, sentences, phrases, and words that are translations of each other.

(ii) Building a translation model based on statistical model, which describes the process of transfer of words, phrases, or grammatical structures, both syntactic and semantic, from one language to another.

(iii) Building a language model, which describes the structure of a statistical model of how likely various sequences of words or grammatical constructions are in the target language.

This model relies on parallel corpora to compute statistics on the frequency with which source and target language words co-occur in aligned segments. The resulting models furnish no absolute translation correspondences but only more or less probable ones. In other words, every target word in the training corpus is viewed as a potential translation of each source word. The mathematics of SMT describes the definition of a series of five increasingly complex translation models that consider an expanding range of factors in establishing the alignments between words. The core to all these models is the bilingual lexicon that is the table that lists the targets equivalents for each source word, along with their probabilities. There are several ways of refining the basic methodology so that the extracted translation lexicons become sufficiently accurate to allow for their integration into the workflow of conventional MT systems. But the main drawback of these models is that they can provide the possible translations of a given source word but do not explain the various contextual factors which determine when a given equivalent is appropriate.

Current MT researches are mostly in the field of high precision in chunk (sentence) and word alignments. The translation model based on the structures of the two languages and their relationship, is obtained after aligning the parallel text down to the level of words and phrases. The simplest models are lexical. Since words are related to the neighbouring words, more refined models generally define their probability distributions over contiguous phrases. Bilingual dictionaries are also a good source for directly mining word correspondences. In addition to aligning individual words and contiguous strings or 'phrases', the attractive approach is to align higher level of syntactic or semantic structures.

The statistical technique in machine translation has led to tremendous increase in the accuracy of translation and quality of research systems. With passage of time, quality of machine translation is improving and its accuracy is being enhanced with the increase in available parallel bilingual data. One of the important factors in these improvements is definitely the availability of large amount of data for training statistical models. Yet the modelling, training, and search methods have also improved since the field of statistical machine translation was pioneered by IBM in the late 1080s and early 1990s.

2. MOTIVATION

Although expert systems incorporate deep linguistic knowledge, these are expensive to maintain and implement on new language pairs. These give only one target sentence for each source sentence, whereas SMT yields many target sentences, each with a probabilistic score. The SMT is not confined to any specific pair of languages. The SMT systems can be trained over the parallel corpus within days to produce the translation.

Using Moses one can extend the phrase-based translation with factors and confusion network decoding. Confusion Network Decoding provides the translation of ambiguous input as well. The Moses decoder is a drop in replacement of Pharoh, the popular phrase-based decoder and uses Bean Search algorithm.

Machine translation (MT) is the computer-based translation from one natural language into another language using computers.[7-8] SMT is an approach to MT that is characterised by the use of machine learning methods. In order to accomplish the task of translation, we have used the technique of SMT was used as it is

accurate on a large corpora and can be used for any language pair. A translation model trained on 700,000 sentences of the German-English Europarl corpus and the open source Moses toolkit were used to perform the factored phrase-based translation.

3. RELATED WORK

The rule-based translation system[9] used in machine translation, should be provided with all the necessary linguistic rules as well as the exceptions to these rules. Thus, rule-based system proves to be time-consuming and costly. The scope of example-based machine translation (EMBT) is quite limited because inspite of the large corpus, everything that one wants to translate cannot be covered by examples.[10] Although much research has been carried out on SMT since 1980, the use of phrase translation has shown improved results. Phrase-based machine translation can be referred to Och's alignment template model.[11] This was further enhanced by Yamada using phrase translation in a syntax-based model.[12] Marcu introduced a joint probability model for phrase translation.[13] Today, phrase-based translation is being used in most of the SMT systems. The phrase-based SMT was defined by Koehn, Och, and Marcu.[14] According to the Factored Translation Model[15], each word is considered along with additional annotations. Factored Translation Model, along with SMT toolkit Moses, is considered to be the state-of-the-art SMT system.[16] The greatest advantage with this model is that it is not built specifically for a language pair and can be used for translation between any two languages, as long as sufficient parallel corpus is available to train the models.[17] Much work has been done in sentence alignment for automatically producing bilingual corpora.[18-21]

3.1 Background

Statistical machine translation is based on finding the most probable translation of a sentence using data gathered from a bilingual corpus. MT is the automatic translation of text from one natural language to another. To perform the task of translation the computer needs to understand the following:

(i) Grammar of both the languages.

(ii) Morphology of setences.

(iii) Syntax.

(iv) Semantics.

(v) Lexicon.

(vi) Pregmatics.

The computer can learn these automatically from parallel bilingual text. The objective of SMT is to extract general translation rules from a given corpus consisting of sufficient number of sentence pairs which are aligned to each other. Alignment of the words of the source language with those of the target language forms the core building block of the translation models. There can be more than one alignment for two sentences that are the translations to each other.

One can express the problem of translating a German sentence *g* into English sentence *e* by the following application of Bayes Rule:

$$\text{argmax } Pr(e|g) = \frac{Pr(e)Pr(g/e)}{\Pr(g)} \quad (1)$$

$$= Pr(e)\ Pr(g/e) \quad (2)$$

where *Pr* (*e*) is the language model probability, and where *Pr* (*g*/*e*) is the translation model probability

One considers all possible English sentences e and choose the one that maximises the product *Pr*(e) *Pr* (*g*/*e*). The factor *Pr* (*g*) can be ignored because it is the same for every '*e*'[22]. One of the ways to accomplish the task of Statistical Machine Translation is to use phrase-based translation.

Phrase-based SMT Model in SMT has enhanced the performance of MT systems. The basis of phrase-based translation is to fragment the input sentence into phrases and then translate these phrases into the target language. These translated phrases in the translated language are then re-ordered. The phrase-based model greatly depends on the size of the parallel corpus.

In contrast to the phrase-based translation model, in Factored Translation Model[23], the surface form is augmented with factors such as root, part-of-speech tag or morphological information.[24] The string is then represented as a vector of strings and a phrase is represented as a sequence of vectors. The translation process is broken up into the following three mapping steps[25] as shown in Fig. 1.

- Translate input lemmas into output lemmas.
- Translate morphological and POS factors.
- Generate surface forms given the lemma and linguistic factors.

The translation model finds out the correspondence between the source sentence and the target sentence and tells whether the two translations match with each other. The translation table maps foreign phrases to the target language. The translation table needs to be extracted from the parallel corpus by the process of word alignment.

The Moses decoder implements the Beam Search (BS) algorithm to find the best translation for the given

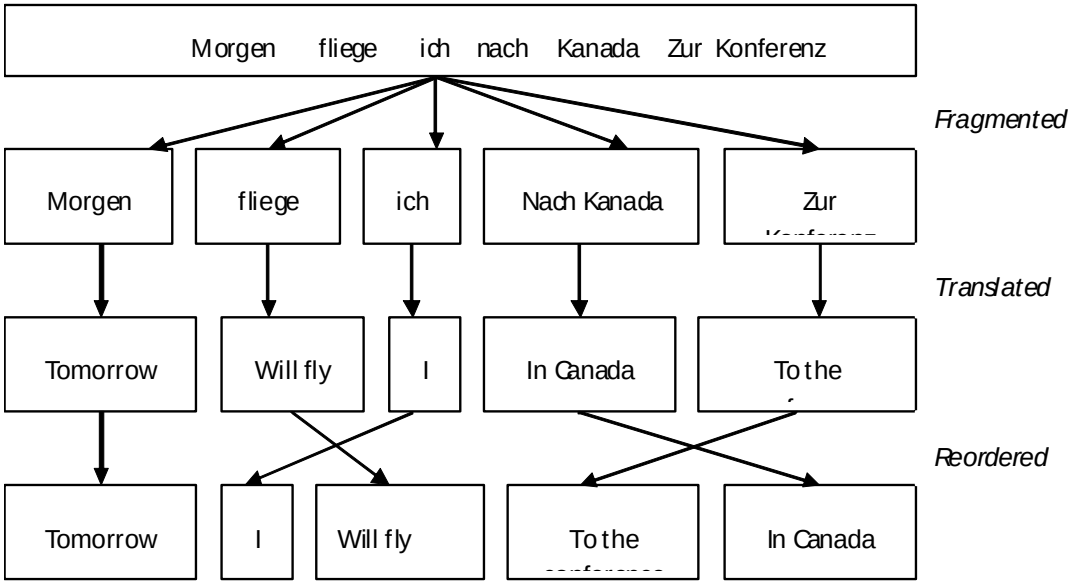

Figure 1. Steps involved in phrase-based translation.

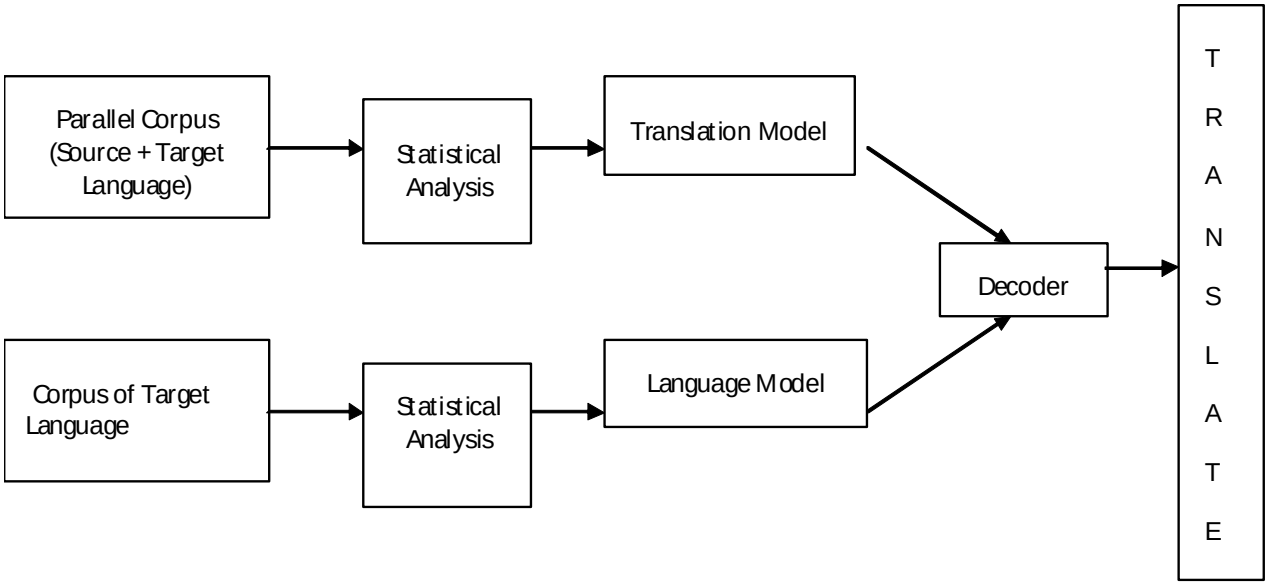

Figure 2. Building blocks of SMT system (phrase-based).

input. One can have a number of phrase translation options for a given input string of words. The BS algorithm searches through all the translation options and selects the best translation having the highest probability.

Confusion network decoding, an eminent characteristic of Moses, allows a network of different word choices to be examined by the machine translation system. It is a blend of speech recognition and machine translation. Confusion network is a linear directed graph[26] and is used to reduce the complexity of the lattices to a series of confusion sets as shown in Fig. 2. Various operations such as sentence scoring, perplexity computation, sentences generation, and various types of model interpolation are performed with N-gram-based and related language models[27]. The language model, build in the target language, decides whether the translation made is in meaningful English language or not.

4. RESOURCES

For SMT the open source toolkit Moses has been used. The English-German parallel corpus has been obtained from the European Parliament Proceedings EROPARL.

4.1 Moses

The Moses translation toolkit has been used to build a machine translation system from one language to another. Moses is capable of learning translation tables, language models, and decoding parameters from the

data. It is released under the LGPL licence and is available both as source code and as binaries for Windows and Linux. It is written in C++ and Perl. The key features of Moses include pre-processing the data, training and tuning of language model, and the translation model. Moses uses the following tools:

- GIZA++: used for word alignment.[28]
- SRILM : for building language model.[29]
- MKCLS: used for training word classes used in SMT.[30]
- NIST, BLEU: for evaluation of the translation.[31-33]

4.2 Parallel Corpus

The corpus for English and German languages were obtained from Europarl parallel corpus, the Proceedings of the European Parliament. It includes versions in 11 European languages.

4.3 Hardware Requirement

The hardware requirements for the SMT are:

- High-end server with the minimum 2 GB RAM Windows/Linux OS
- Operating System Used: Red Hat Enterprise Linux 5.1.19.6

5. METHODOLOGY

The required tools GIZA++, SRILM, MKCLS, NIST, BLEU were downloaded and successfully installed along with the required scripts. The given parallel corpus was preprocessed. The English and German sentences were aligned, concatenated and then tokenised. Long sentences were first filtered and then lowercased.

The language model was then built in the target language, i.e, in English. For this, the English language model data was first tokenised and then lowercased and the language model was built using SRILM. A 5 gram language model was built. The phrase-based translation model consists of the following two files:

- Phrase-table: the phrase translation table.
- Moses.ini: the configuration file for the decoder.

Once the language model was built, the training process was started. The training process was executed by the script *train-factored-phrase-model.perl.* In the training process, various steps were involved. First the training data was prepared and then the vocabulary files of both the languages were generated and the parallel corpus was converted into a numbered format. Second, the words were aligned bidirectional. The maximum likelihood lexical translation table can be obtained from the aligned words. Next, the phrases were extracted and piled up into one big file. A translation table was created from the stored phrase translation pairs. The phrase translation probability distribution for the German phrase and the English phrase was calculated, which helped in computing the translation scores. The lexicalised reordering model gives the cost which is linear to the reordering distance. Finally, the configuration file for the decoder, called *Moses.ini*, was generated. The training process terminates here and the translation in the target language can be obtained for any given input sentence in the source language. The translated sentences were evaluated using the machine translation scores BLEU, and NIST. GIZA++[28] which uses the IBM Model 1,2,3 and 4 with 5,0,3,3 iterations, respectively.

The model for translating German text into English was trained. The input given (in German) was echo 'wir wissen nicht,was passiert' | / usr / home / smt / projects / programming / language_model / Moses/ Moses-cmd / src / Moses –f Moses.ini > out.txt

The result obtained was

```
Translating: Wir wissen nicht , was passiert
Collecting options took 0.010 seconds
Search took 0.170 seconds
BEST TRANSLATION: Wir do not know what is happening
Translation took 0.180 seconds
Finished translating
```

6. RELATED ISSUES

The major issue concerned with SMT is the heavy resource requirement while handling large corpus. A large amount of time is spent in alignment of the corpus and to train the corpus. Moreover, high processing power is required. Parallel corpus of sufficient size is not available for every language and the corpus is domain specific. Comparable non-parallel corpora[34] can be used to overcome the limitation. Parallel sentences can be identified and extracted from very large comparable corpora of newspaper articles.

7. FUTURE WORK

New developments in machine learning are being applied to MT also. More linguistic knowledge can be introduced in the models and in parametric estimation. Syntactic modelling is also an area of active research. Modelling techniques and parameter estimation methods are also subject to refinement.[35] In future, advancement of SMT lies in integrating SMT with speech recognition, web page translation, information retrieval, mining parallel corpus from the World Wide Web[36], spoken language

understanding[23], document summarization and other Neuro-linguistic programming applications.

One can improve the translation process in many ways. We have used similar sentence retrieval and weighting schemes for translation purpose which are very simple. It may work better by extracting a large dictionary and using semantic information to generalise the training data.[37] One may introduce a new strategy for integrating WSD into an SMT system that performs fully phrasal multi-word disambiguation.[38] Integration of Named Entity Recognition and a transliteration system with Moses might further improve accuracy of translation as any named entity can be directly recognised and transliterated.[39]

8. CONCLUSION

A statistical translation model based on bilingual phrases has been presented. Use of bilingual phrases instead of single words in the translation model has significantly improved translation quality. The task of translating the German text into English was accomplished by SMT using Moses. The translated output gives more accurate translation if the model is trained repeatedly and intensively, with a larger corpus containing more number of sentence pairs.

REFERENCES

1. Chomsky, N. Three models for the description of language. *IRE Transactions on Information Theory*, 1956, **2**(2),113-24.
2. Chomsky, N. Formal properties of grammars. Handbook of Mathematical Psychol. 1963, John Wiley and Sons, New York, **2**, 323-418.
3. Chomsky, N. & Miller, G. Finite-state languages. *Inform and Control,* 1958, **1**, 91-112.
4. Koehn P Europarl. A parallel corpus for statistical machine translation. *In* MT Summit X, the Tenth Machine Translation Summit, Phuket, Thailand, 2005. pp. 79-86
5. European Parliament Proceedings-Parallel Corpus 1996-2006. http://www.statmt.org/europarl/
6. Berger, A. L.; Brown, P. F.; Della Pietra, S. A.; Della Pietra, V. J.; Gillett, J. R.; Lafferty, J. D.; Mercer, R. L.; Printz, H. & Ureš, L. The candide system for machine translation. *In* Workshop on Human Language Technology, edited by NJ Plainsboro, 08–11 March 1994. Paper presented at Human Language Technology Conference. Association for Computational Linguistics, Morristown, NJ. pp. 157-62. http://dx.doi.org/10.3115/1075812.1075844
7. Goswami, Sumit; Shah, Nirav; Roy, Devshri & Sarkar, Sudeshna. *In* 6th International Conference on Natural Language Processing, ICON-2008:Pune NLP Tools Contest: Statistical Machine Translation (English to Hindi).
8. Shah, Nirav & Goswami, Sumit. Statistical machine translation using Moses. *PC Quest,* 2009,
9. Hutchins, John. Trends in machine translation research. University of East Anglia, Norwich, England, http://www.hutchinsweb.me.uk/SprogErhverv-1995.pdf (accessed on 13 December 2009).
10. Phillips, Aaron B.; Cavalli-Sforza, Violetta & Brown, Ralf D. Improving example-based machine translation through morphological generalisation and adaptation. *In* Machine Translation Summit XI, Copenhagen, Denmark, September 2007.
11. Och, F.J. & Ney, H. The alignment template approach to statistical machine translation. *Computational Linguist*ics, 2004, **30**(4), 417-49.
12. Yamada, K. & Knight, K. A syntax-based statistical translation model. *In* 39th Annual Meeting on Association for Computational Linguistics, Toulouse, France, 06–11 July 2001. Annual Meeting of the ACL. Association for Computer Linguistics, Morristown, NJ, 523-30.
13. Marcu, D. & Wong, W. A phrase-based joint probability model for statistical machine translation. *In* Conference on Empirical Methods in Natural Language Processing, ACL-02, Volume 10, Morristown, NJ, 2002. pp. 133-39.
14. Koehn, P.; Och, F. J. & Marcu, D. Statistical phrase-based translation. *In* North American Chapter of the Association for Computational Linguistics on Human Language Technology, Edmonton, Canada, May 27-June 01, 2003, Morristown, NJ, 2003. pp. 48-54.
15. Koehn, P.; Hoang, H.; Birch, A.; Callison-Burch, C.; Federico, M.; Bertoldi, N.; Cowan, B.; Shen, W.; Moran, C.; Zens, R.; Dyer, C.; Bojar, O.; Constantin, A. & Herbst, E. Moses: Open source toolkit for statistical machine translation. *In* 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, Prague, Czech Republic, 25-27 June 2007). NJ, 2007. pp. 177-80.
16. Tufis, Dan & Ceausu, Alexandru. Factored phrase-based statistical machine translation. Research Institute for Artificial Intelligence, Bucharest, Romania.
17. Finch, A. & Sumita, E. Using statistical machine translation to grade training data. Language Translation Group, Kyoto, Japan.

18. Chen, S. F. Aligning sentences in bilingual corpora using lexical information, *In* 31st Annual Meeting on Association for Computational Linguistics, Morristown, NJ, USA, 1993, pp. 9-16.

19. M. Kay & M. R¨oscheisen. Text-translation alignment. Computational Linguistics, 1993, **19**(1), 121-42.

20. Fattah, M. A.; Bracewell, D.B.; Ren, F. & Kuroiwa, S. Sentence alignment using P-NNT and GMM. *Comput. Speech Lang*., 2007, **21**(4), 594-608.

21. Utsuro, T. Ikeda, H.; Yamane, M.; Matsumoto, Y. & Nagao, M. Bilingual text matching using bilingual dictionary and statistics. *In* 15th Conference on Computational Linguistics, Morristown, NJ, USA, 1994, pp. 1076-82.

22. Brown, Peter E.; Stephen, A.; Della, P.; Vincent, J.; Pietra, Della; Mercer, Robert L. The mathematics of statistical machine translation: Parameter Estimation. Computational Linguistics, 1993, **19**(2), 263-311.

23. Klaus, Macherey; Bender, Oliver & Ney, Hermann. Applications of statisticalmachine translation approaches to spoken language understanding.

24. Federico, Marcello; Shen, Wade. Open source toolkit for statistical machine translation: Factored translation model and confusion network decoding. *In* Language Engineering Workshop, 2006. http://www.statmt.org/jhuws/ (accessed on 13 December 2009).

25. Moses: Statistical machine translation system user manual and code guide. http://www.statmt.org/moses/manual/manual.pdf

26. Allauzen, A. Error detection in confusion networks. Université Paris-Sud, France., ftp://tlp.limsi.fr/public/allauzen_is07.pdf. (accessed on 13 December 2009).

27. N-gram language model. http://www.speech.sri.com/projects/srilm/manpages/ngram.1.html (accessed on 11 December 2009).

28. GIZA++: Training of statistical translation models, http://www.fjoch.com/GIZA++.html (accessed on 11 December 2009).

29. Downloading and building SRILM. http://www.speech.sri.com/projects/srilm/download.html (accessed on 11 December 2009).

30. MKCLS: Training of word classes. http://www.fjoch.com/mkcls.html (accessed on 11 December 2009).

31. Papineni, K.; Roukos, S.; Ward, T. & Zhu, W. BLEU: A method for automatic evaluation of machine translation. *In* 40th Annual Meeting on Association For Computational Linguistics, Philadelphia, Pennsylvania, 07-12 July 2002, Morristown, NJ. pp. 311-318.

32. Turchi, M.; De Bie, T. & Cristianini, N. Learning performance of a machine translation system: A statistical and computational analysis. *In* 3rd Workshop on Statistical Machine Translation, Columbus, Ohio, 19 June 2008, Morristown, NJ, pp. 35-43.

33. BLEU Evolution script. ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-v11b.pl (accessed on 11 December 2009).

34. Munteanu, Dragos Stefan & Marcu, Daniel. Improving machine translation performance by exploiting non-parallel corpora. Computational Linguistics, 2006, **31**(4), 477-504

35. Lopez, A. Statistical machine translation. University of Edinburgh. http://homepages.inf.ed.ac.uk/alopez/papers/survey.pdf (accessed on 14 December 2009).

36. Li, B.; Liu, J. & Shi, W. Web-based parallel corpora for statistical machine translation. *In* 6th International Conference on Machine Learning and Applications, 13-15 December 2007, ICMLA, IEEE Computer Society, Washington, DC. pp. 444-49. http://dx.doi.org/10.1109/ICMLA.2007.117

37. Matthias, Eck; Vogel, Stephan & Waibel, Alex. Improving statistical machine translation in the medical domain using the unified medical language system, *In* 20th International Conference on Computational Linguistics, Geneva, Switzerland, 2004.

38. Marine, Carpuat & Dekai, WU. Improving statistical machine translation using word sense disambiguation. *In* Joint conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, June 2007. pp. 61–72.

39. Babych, Bogdan & Hartley, Anthony. Improving machine translation quality with automatic named entity recognition. *In* 7th International EAMT Workshop on MT and Other Language Technology Tools, EACL 2003: 10th Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary.

About the Authors

Shri Mukesh holds MCA from IGNOU, Delhi and MSc (Computer Science) from MDU Rohtak. He has more than 10 years of experience in software development, website designing and hosting, Linux, Windows, MySQL, Visual Basic, and PHP.

Shri GL Vatsa holds postgraduate degree in linguistic and degree in German and French. He has been working as translator in DESIDOC for the last 27 years. He is responsible for translating documents from European language, viz., French and Dutch to English and also has experience in translating Japanese and Chinese literature using machine translation.

Ms Nikita Joshi is an MTech (Computer Science and Engineering) student at Amity University, Uttar Pradesh. Her research interests include machine translation, data mining, ensemble learning, natural language processing, and wireless communication. Her recent research work has been on Impact of Lambda on Bagging. She has extensive experience on Linux, Windows, Oracle, Java, C++, C and ASP.NET.

Shri Sumit Goswami holds MTech in Computer Science and Engineering from IIT Kharagpur, Post Graduate Diploma in Journalism and Mass Communication, Bachelors degree in Library and Information Science, and BTech in Computer Science and Engineering. He has been working as a Scientist with Defence Research and Development Organisation (DRDO), New Delhi since 2000. His areas of interest include network centric operations, mobile ad hoc and sensor networks, web-hosting security, text mining and machine learning. He has published 41 papers/chapters in various journals, books, conferences and seminars. Sumit also chaired a session on the theme Computer Architecture in National Conference on Emerging Principle and Practice of Computer Science held at GND Engineering College, Ludhiana, India in August 2006.