International Journal of Researches in Biosciences, Agriculture & Technology | January 2014 Issue-2, Volume-1 | ISSN No. (Online) 2347-517X

ANALYTICAL STUDY OF MACHINE TRANSLATING SYSTEM IN NLP

*[1]**P. V. Nimbalkar, P. K. Butey**
[1]Dr. Ambedkar college,Nagpur-10
[2]Kamla Nehru Mahavidyalaya, Nagpur-27
E-mail : pnimbalkar20@rediffmail.com

Abstract

Automatic methods of measuring MT performance give us the ability to evaluate various features and present objective results without worrying about agreement rates of humans[1]. The problem of human agreement has been shifted to the set of references that are developed for a set of documents and each test system can then be evaluated objectively since the set of references are common to all systems[2].

Keywords: Machine Translation, Natural language processing.

1 Introduction

Human languages have evolved over a significant period of time and although initially in journey as a species it may have been sufficient to know the local language ("mother tongue") and perhaps the language of neighboring peoples, current trends dictate that we are able to process material produced in many languages[3]. The Web and global access allows for access of content; however, comprehending the content requires a multilingual person or for the majority of people, an automated way of rendering the content into the user's preferred language is required. In building MT models, there are two major problems that need to be addressed:

• Word Order: Translation is normally done at the sentence level and it might very well be that the last token in the source sentence is the key informant to the first token in the target sentence.

• Word Choice: Each source token can be represented in the target language in a variety of ways.

These two problems are not independent and the order in which translate the source tokens directly affects which words might be used in the output sentence. As an example from an Arabic–English MT test

(NIST, 2003), Machine Translation can be viewed as taking the source sequence *S* and performing increasing amounts of analysis as suggested by the pyramid (Vauquois, 1968). At the base of the pyramid, words can be transferred from the source to target language. As we go up the pyramid, the level of sophistication increases and at the very top[4].

2 Approaches

Warren Weaver's memorandum (Weaver, 1955) clearly initiated ideas in the statistical approach to MT;however, it was the pioneering work of the IBM group (Brown et al., 1993) in the early 1990s that led to the renewed and sustained interest in the statistical approach to MT. While initial efforts in SMT were mostly word-based. In Addition, natural language parsers have been developed and this has led to both Syntax and Hierarchical-based approaches. The early measures of MT included 'Adequacy' and 'Fluency' and these utilized human evaluators and may be defined as follows[5].

Adequacy. Does the translation capture an adequate amount of the meaning of the sentence in

the source language?

Fluency. Is the translation fluent in English?

these two measures. For a test set, a few references are developed by asking independent translators to provide translations. BLEU then computes the precision of various length strings and the final score is weighted sum with a penalty if the system produces a very short translation. This penalty acts as a measure of the recall of a system. BLEU revolutionized the rate of progress of statistical systems since it made possible to run many evaluations automatically. The list of MT metrics utilized currently is quite long and dedicated to finding better metrics; the major alternatives are now: (a) Translation Error Rate (TER) (Snover et al., 2006), (b) METEOR (Banerjee and Lavie, 2005).Most major MT evaluations in addition to automatic methods, utilize human editors to edit the system outputs and compute TER of the system output

relative to the edited string, which is termed Human-TER or HTER.Automatic methods of measuring MT performance gives us the ability to evaluate various features and present objective results without worrying about agreement rates of humans. The problem of human agreement has been shifted to the set of references that are developed for a set of documents and each test system can then be evaluated objectively since the set of references are common to all systems[6].

3.Translation Models

In the original formulation of Brown et al. (1993), the source-channel model has a component, $p(s|t)$, which involves both the source and the target languages and is named the "translation"
model.*)*.

A conditional probability model can be expanded using the normal chain rule as,

$p(S|T) = p([\ s0\ s1\ \ldots\ sk\ ]|[\ t0\ t1\ \ldots\ tl\ ])$

$= p(s0|[\ t0\ t1\ \ldots\ tl\ ])p(s1|s0, [\ t0\ t1\ \ldots\ tl\ ])\ \ldots$

3.1 Phrase-Based Systems

The order of applying the alignment templates can be represented using a hidden variable. In this approach , the hidden variables in the process are (a) the optimal segmentation of the source, (b) the order of applying the alignment templates. The search process yields the optimal values for these variables subject to the limits of the search. The alignment template remembers the internal alignment of the words and this is used in the context features used in that system. Other phrase-based approaches often drop the alignment and compute the features for the phrase-pair when the phrase-pair is created. The typical features in a phrase-based system have significant overlap with Och et al. (2004) and they will be used in subsequent systems. Perhaps the most attractive feature of the AT approach is the extensibility of the feature set and the ability to train parameters for arbitrary features. The maximum entropy

formulation (Della Pietra et al.,1995) was suggested in Papineni et al. (1997, 1998) as a model for natural language understanding. In MT, Och and Ney (2002) utilize the GIS algorithm for training the parameters of the AT. Three issues arise in the training of this model (Och and Ney, 2002)[7]:

• Normalization: The normalization *of* the exponential model requires the sum over many target

Sequences. Obviously, some approximation is required and in the AT system the sentences being

Summed over in the denominator are the very probably sentences derived from an n-best algorithm.

• Multiple-references: Unlike speech recognition and indeed many pattern recognition problems,

the translation of a source sentence is ambiguous and there are many ways to render the meaning

in the target language. In the AT system, the optimization criterion is modified to reflect many

references.

• Reachability: Occasionally, the n-best is insufficient and the references are not in the n-best list.

This problem is solved by selecting as the reference translation those that have the minimal number of word errors given the reference translations. Despite these problems, phrase-based systems are the workhorse of SMT systems due to their simple and relatively straight forward method of extracting phrase libraries and training weights. Systems for new language pairs that have parallel corpora can be built by utilizing the GIZA++ toolkit for generating word-alignments and the open source phrase decoders such as Moses (Koehn et al., 2007).

3.2 Syntax-Based Systems for Machine Translation

Syntax is the study of the grammar of a language and in particular how phrases and clauses are put together. Syntax-based approaches rely on

parsing the source, or the target, or in some cases both languages. The parses are obtained from a statistical parser trained for each language; for English, the parser is trained on the Penn Treebank (Marcus et al., 1993)[8] and the Arabic parser on the Arabic Treebank (Bies and Maamouri,2008). Parsing output can be either as a Constituent Parse or a Dependency Parse. A constituent parse is a rooted tree whose leaves are the original words of the sentence. The internal nodes of the tree cover a contiguous sequence of the words in the sentence (usually called a span) and to each of these internal nodes is associated a label that describes the syntactic role of the words under this node. A dependency parse shows for each word in the sentence the "parent" or "head word." Consider the English phrase "local time" and its Arabic translation "Altwqyt AlmHly"; in both languages, they have a NP node that spans just these phrases. Syntax-based systems could learn to reorder the words for the translation process by collecting these tree fragments with respect to the alignments. In Yamada and Knight (2001), decomposition similar to the IBM models is developed for a syntax-based approach that transforms the target parse tree using the operations of insertion, reorder, and translation. An efficient graph representation of these operations allows the model to compute the required parameters. Translations often have to break the parse structure and this has been studied in Fox (2002) and a better method of obtaining rules is discussed in Galley et al. (2004). From the original MT pyramid, we might expect that syntax-based systems should be more general or at least less susceptible to sparsity caused by finite training data; however, the challenge seems to be in reducing the errors in the parses and part-of-speech taggers. Source sentences that are difficult to translate seem to also be a challenge for

Parsers[9].

3.3 Direct Translation Models

The noisy-channel model was motivated in Brown et al. (1993) on the observation that the source sentence is well-formed and that the MT hypotheses can generate arbitrary strings which are often not well-formed. An equivalent approach for SMT is the direct model where we combine log-linearly various models. The inventory of phrases in such systems is highly redundant which shows a set of phrases that cover the two-word Arabic fragment “mrAkz AlAqtrAE” The unigram count of a word is greater than or equal to any bigram count that involves the same word. Counts are equal only in the rare case of a completely sticky pair (perhaps “Humpty Dumpty” in English) or with words that occur very few times due to data

sparseness.In DTM2, a minimalist set of phrases is used that are prescribed by the word alignments. In contrast,Blunsom et al. (2008) discusses a global method that allows overlapping phrases[10].

4. Features

The direct model approach shares the advantage of an extensible feature set with the AT (Automatic Translation) approach. Generalized Iterative Scaling (Darroch and Ratcliff, 1972), Improved Iterative Scaling (IIS) (Della Pietra et al., 1995), and a variety of conjugate gradient methods. In Ittycheriah and Roukos (2007), In general, the features are language neutral and the only exception is the segmentation feature that varies from characters for Chinese to

prefix, suffix, and stems for most other languages[11].

4.1Search Strategies

Above methods to incorporate linguistic notions as features and training strategies for MT.The search problem has been addressed for speech recognition; for phrase-based MT was developed in Tillmann and Ney (2003), shown to be NP-complete for Model 1 type decoders (Knight, 1999), and an *A** algorithm for the AT approach for MT in Och and Ney

(2004). Syntax-based approaches (e.g., Chiang, 2005; Zollmann and Venugopal, 2006) utilize CKY parser (Earley, 1970) and extensions for MT. For brevity, we sketch out the decoding strategy used in the DTM approach above but recognize that heuristic search is the subject of much ongoing research[12].

In DTM, a beam-search algorithm is employed that is quite similar to the AT approach. The search process proceeds from left to right in a source sentence considering a *window* of source positions at each time. A second parameter, *skip*, controls how many source positions in the window can be in an "open" state. DTM allows translations that contain a variable and this requires the state of a source position to be in one of three conditions: {open, partially covered, covered}. The output of the beam search is a lattice, which has all the hypotheses that have been explored during the search. The best hypothesis at the end can then be back-tracked to produce the translation output. Each hypothesis captures the following:

• *ti*, the target production at this step.

• *lmi*, the language model state.

• *Ci*, the coverage status (implemented as two binary coverage vectors to capture partial and covered notions).

• *partiali*, the data required to complete a state that is in partial state: (a) an index to the next target to be produced for the partial phrase, (b) the source position that is in partial state[13]. This information can be recovered from the above coverage status but for efficiency we store the extra data.

• *scorei*, the score of this hypothesis.

• Back-pointer to do the final trace back through the lattice.

At each extension the following steps are carried out:

• ComputeSourcePositions: This function returns a vector of source positions that should be

considered for generating at this point.

• Extend: This function generates a set of hypothesis from the phrase pairs aligned to the source
positions obtained in the above step. This step also computes the weighted combination of scores
from (a) the translation models, (b) the language models, (c) a word-count score. The translation
models include the direct model and the model 1 score (Equation 17.10) for each direction. Since
the language model cost increases with the number of words produced, the word-count score
encourages the system to produce longer phrases.

• Merge: Paths are merged and the best one is kept when the language model state, the coverage
status, and the current production are identical[14].

• Prune: Paths are pruned to keep, *k*, hypothesis for each coverage pattern. When all alive hypotheses have no more open source positions, the search is terminated and the best sequence is output.

5. Conclusion

The MT pyramid seems to offer the best roadmap and methods are being sought to
inject more complex information into the translation systems. Constituent and dependency parsers as well as Named Entity taggers for various languages are now available and translation systems should be able to improve their performance by incorporating these types of information into the systems. Current systems are sentence oriented and this focus has left many document-level effects unattended including (certainly not an exhaustive list): (a) pronouns (when to use them, which one to use, etc.), (b) articles (e.g., in English to determine the definiteness of a noun would require document-level analysis), and (c) tense. Algorithms and methods for more robust estimation of translation

models from parallel corpora that are automatically collected is being actively pursued.

7.References

Banerjee, S. and Lavie, A. (2005). Meteor: An automatic metric for MT evaluation with improved correlation withhumanjudgments. In *Proceedings of Workshop on Intrinsic and Entrinsic Evaluation Measures forMTand Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL-2005)*, Ann Arbor, MI, pp. 65–72.

Bies, A. and Maamouri, M. (2008). Arabic treebank morphological and syntacticzannotation guidelines.http://projects.ldc.upenn.edu/ArabicTreebank Block, H. U. (2000). Example-based incremental synchronous interpretation. In Wahlster, W., editor,*Verbmobil: Foundations of Speech-to-Speech Translation*, pages 411–417. Springer-Verlag, Berlin,Germany.

Blunsom, P., Cohn, T., and Osborne, M. (2008). A discriminative latent variable model for statistical machine translation. In *46th Annual Meeting of the ACL*, Columbus, OH.

Brants, T. and Franz, A. (2006). Web 1T 5-gram version 1. *Linguistic Data Consortium*, LDC2006T13,.

Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., and Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

Chiang, D. (2005). A hierarchical phrase-based model for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the ACL*, pp. 263–270, Ann Arbor, MI.

Darroch, J. N. and Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, 43(5):1470–1480.

Della Pietra, S., Della Pietra, V., and Lafferty, J. (1995). Inducing features of random fields. Technical Report, Department of Computer Science, Carnegie-Mellon University, Pittsburg, PA, CMU-CS-95-144.

Earley, J.(1970). Anefficient context-free parsing algorithm. *Communications of the Association for Computing Machinery*, 13(2):94–102.

Fox, H. J. (2002). Phrasal cohesion and statistical machine translation. In *Proceedings of EMNLP-02*,Philadelphia, PA, pp. 304–311.

Galley, M., Hopkins, M., Knight, K., and Marcu, D. (2004). What's in a translation rule? In *HLT-NAACL*,Boston, MA, pp. 273–280.

Ge, N. (2004). Improvement in Word Alignments. *Presentation Given at DARPA/TIDES MT Workshop*,Gaithersburg, MD.

Hovy, E. (1999). Toward finely differentiated evaluation metrics for machine translation. *Proceedings of EAGLES Workshop on Standards and Evaluation*, Pisa, Italy.

Ittycheriah, A. and Roukos, S. (2005). A maximum entropy word aligner for arabic-english machine translation. In *HLT '05: Proceedings of the HLT and EMNLP*, Vancouver, Canada, pp. 89–96.

Ittycheriah, A. and Roukos, S. (2007). Direct translation model 2. In *Human Language Technologies 2007: The Conference of the NA-ACL*, Rochester, NY, pp. 57–64. Association for Computational Linguistics.

Jelinek, F. (1999). *Statistical Methods for Speech Recognition*. MIT Press, Cambridge, MA.

Knight, K. (1999). Decoding complexity in word-replacement translation models. *Computational Linguistics*, (25):607–615.

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., and Herbst, E. (2007). Moses: Open source