ISSN 2229-5518

Statistical Machine Translation

Akanksha Gupta, Shubham Pathak, Kannadasan R

Abstract - We will be talking about Statistical machine Translation as an Aspect of machine learning model. How the SMTs use a heuristics and various models based on probability distribution to translate large chunks of language to another language which is approximately close enough. SMTs generally work by examining the many instances of human-produced translation thus maintaining a heuristic of the historical data set. This research gives a brief summary regarding the various research models that are implemented in Statistical machine translation. We will describe the scope of current research in this field, a formal problem description and various sub problems. Finally, we conclude with an overview of evolutions as well as the future prospects.

Index Terms - Statistical Machine Translation, Natural Language Processing, Bilingual Corpus, Bayes Probability Function, Natural Language

1. INTRODUCTION

As we study the implementation of SMTS, one common result is that a fair amount of the modified code consists of similarities or consistencies which are observed using some statistically based learning model like the machine learning model. An alternative solution to the SMT method is the rule-based method which manually provides all the laws related to translation from one language to another like the semantics used, grammar rules, and knowledge base and environment variables. But such an approach is often vulnerable to changing platforms and also it is difficult to specify all the set of possible rules for machine translation manually.

2. MOTIVATION

Although there are many expert systems which incorporate deep linguistic knowledge based learning strategies, these are expensive to maintain cost wise as well as difficult to implement on new language pairs Also, they use one target sentence for each source sentence whereas SMT gives various different target sentences with a probability for each one. Again SMTs are not confined to a specific pair of languages and they can be trained within days to develop parallel corpus to give the translation

3. RULE BASED APPROACH AND APPROXIMATION SOLUTION

3.1 In association with the rule based approach

The results of SMT are quite satisfactory when we are working with an application program interface or APIs as we can provide translation rules for various different phrases. But if we are given some random set of variables or parameters then our SMTs may give a poor output. To remove this issue we provide our SMTs with heuristics or some prior knowledge of algorithms. For this various training datasets can be deployed which can provide our SMTs, with a good knowledge base. As of now, the traditionally defined rules are being applied i.e. the user provides the syntax. However we also have to find a way to extend such rules so that they can be used for wider application. Refer to Fig. 1.

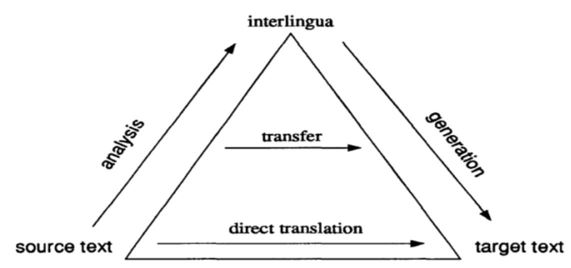

Fig. 1 The rule-based MT system pyramid

3.2 Approximate Solution:

A statistical translational model need not be exact which means that it just has to predict a value keeping the semantics of grammar in mind and that value should be fair enough close to the desired value though not precise. Several synthesizers are designed this way, to predict likely program completion which is then shown to the program analyzer for further scrutiny.

4. ISSUES RELATED TO STATISTICAL MACHINE LANGUAGE PARSING

So far we have seen a direct implementation of SMTs using traditional approaches. One thing to be noted was that most of the programming languages are based on an architecture that is closely based on the grammar syntax of that language. As a result, if such syntax is not taken into account as the programming language evolves our SMTs are bound to produce error- prone translation and consequently the processing time would increase since the system has to deal with debugging of such errors. One way to resolve this is to simply keep a database that records the most recent semantics related to the grammar of the language being translated and then give translations which are based on this grammar. This can highly reduce the errors generated by SMTs. During the decoding process SMTs keep on continuously generating prefixes until the output sentence matches with source sentence, but during this process, it is possible that machine may generate prefixes that do not exist and in such case, it becomes necessary to eliminate those meaningless or extra prefix.

ISSN 2229-5518

To accommodate this issue we can directly define prefix grammar, so whenever a new prefix is generated it will be compared with this prefix grammar to check whether this prefix can be implemented or not, i.e. whether the prefix generated is meaningful or not and based on this observation we can keep or discard the generated prefix. Informally speaking, the prefix grammar is capable of parsing all those strings that can be parsed by the initial string and also the prefixes of those strings. So in SMT, it could be checked for those extra prefixes whether they can be parsed or not.

5. IMPLEMENTATION AND ARCHITECTURE MODEL FOR STATISTICAL MACHINE LANGUAGE TRANSLATION

5.1. Source channel model architecture of statistical machines based on Baye's probability function

Refer to Fig. 2.

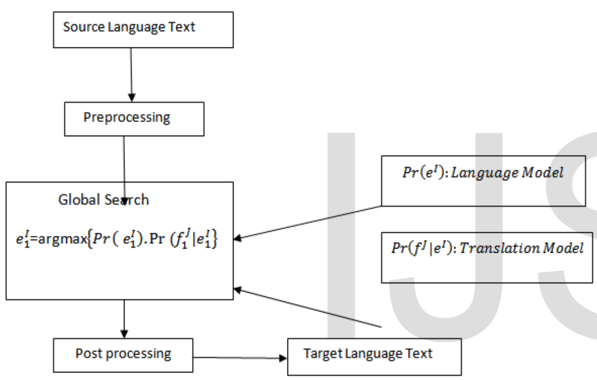

Fig. 2 Architecture diagram using the source channel model

We obtain the following decision rule:

$$e_1^I = \text{argmax}\{Pr(e_1^I).\,Pr(f_1^J|e_1^I)\}$$

Usually the current SMT apply this method. But such kind of translation offers several disadvantages like 1.) The combinational language architecture, as well as the translational language architecture, provides optimal results when true probability distribution is used. However, training models only provide an approximate of the true probability distribution. Therefore we have to use various different distributions to obtain a better result.2.) No straightforward way to extend the SM model by adding additional dependencies.3.) Sometimes we obtain the same result on changing the decision rule. So if such is the case then we have to use that decision rule that gives a more efficient search.

5.2. Direct Maximum Entropy Translation model

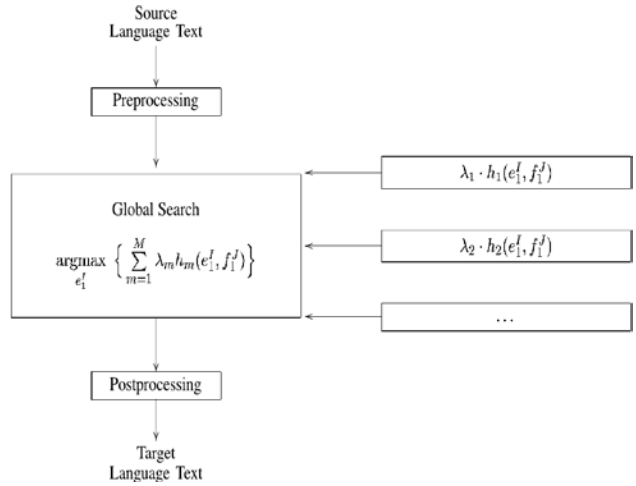

Fig. 3 Model for Direct Maximum Entropy Translation

In order to understand natural language processing, we often implement the above-given strategy. According to this principle, the quantity with the largest entropy is one having the highest probability distribution function based on some given constraints. It's an alternative to source channel model, and further detailed explanation can be found in Berger et al, 1996.s

6. ALIGNMENT MODELS AND MAXIMUM APPROXIMATION

In most of the cases the probability function $Pr(f_1^J|e_1^I)$ is broken down using extra latent variables. Now in the described alignment model $pr(f_1^j, a_1^j\,|\,e_1^i)$, is the hidden or latent variable.

$$Pr(f_1^J|e_1^I) = \sum_{a_1^j} pr(f_1^j, a_1^j|e_1^i)$$

The alignment based modeling is as follows j->i=a_j and then maximum approximation principle is applied from the initiating position j to the final destination i.

So in the final search list we have we all possible combinations of target language sentence and with all possible alignments of sentences. We can also show the dependence on hidden variables using extended feature function.

7. CONCLUSION

In this discussion we saw various different models for statistical machine translation like the source channel architecture model which is a more popularly used architecture nowadays. It uses a naïve Baye's classifier function to generate the probability function for different text sentences. It also supports the basic translation systems which can be expanded by adding some extra functions, so the prediction is more accurate. This is one of the interpretations of statistical machine translation systems as an extension towards bayes decision rule. We saw another interpretation in the form direct maximum entropy model by adding a feature function. Several aspects of dynamic programming can also be included to handle the complex features of entropy-based model. This approach can be quite useful in error reduction as well as optimization of the output of machine translation. Several applications of these exist in the field of pattern recognitions.

REFERENCES

[1]. J. Peters and D. Klakow, Compact maximum entropy language models In Proc. of the IEEEWorkshop on Automatic Speech Recognition and Understanding,1999.

[2]. K. A. Papineni, S. Roukos, and R. T. Ward. Maximum likelihood and discriminative training of direct translation models. In Proc. Int. Conf. on Acoustics,Speech, and Signal Processing, pages 189–192, 1998.

[3]. W. Wahlster Verbmobil: Translation of face-to face dialogs In Proc. of MT Summit IV, pages 127–135, Kobe, Japan, July, 1993.

[4]. R. L. Mercer The mathematics of statistical Machine translation: Parameter estimation, 1993

[5]. P. Beyerlein, Discriminative model combination In Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding, pages 238–245, Santa Barbara, CA, December,1997.

[6]. L. R. Bahl, P. F. Brown, P. V. de Souza, and R. L. Mercer, Maximum mutual information estimation of hidden markov model parameters,1986.

[7]. B. H. Juang, W. Chou, and C. H. Lee, Statistical and discriminative methods for speech recognition in A. J. R. Ayuso and J. M. L. Soler, editors, Speech Recognition and Coding - New Advances and Trends.Springer Verlag, Berlin, Germany,1995

[8]. H. Ney, on the probabilistic-interpretation of neural-network classifiers and discriminative training criteria. IEEE Trans. on Pattern Analysis and Machine intelligence, 17(2):107–119, February, 1995.

[9]. Ayan,N.F and Dorr,B, Going beyond AER. An extensive analysis of word alignments and their impact on MT in proceedings of ACL-COOLING 9-16, 2006.

[10]. Ayan,N.F and Dorr,B and Monz, Alignment link projection using transformation-based learning in Proceedings of HLT-EMNLP.185-192, c.2005a

[11]. Ayan,N.F and Dorr,B and Monz, Neuralign :Combining word alignments using neural networks. In proceedings of HLT-EMNLP.65-72, c.2005b

[12]. Chelba,C and Jelinex F, Exploiting syntactic structure of language modeling in proceedings of ACL-COOLING.225-231, 1998

[13]. Chan,Y.S,NG,H.T and Chiang,Word sense disambiguation improves statistical machine translation in proceedings of MT summit IX, D.2007.

[14]. Dorr, B.J, Jordan, P.W And Benoit, J.W, A survey of current paradigms in machine translation. In advances in computers, M.Zelokowitz, Ed Vol.49.Academic Press, 1-68,1999.

[15]. Fraser, A. and Marcu, D.2006.Semi-supervised training for statistical word alignment. In proceedings of the association for computational linguistics (ACL).769-776.

[16]. Knight,K., Automating knowledge acquisition for machine translation, AI Mag.18,4,81-96,1997.

[17]. LiangP.,Taskar B and Klein D, Alignment by agreement in proceedings of HLT-NAACL.41-48, 2006

[18]. Manning C.D and Schutze, Foundations of Statistical Language Processing.MIT press, Cambridge, MA, H.1999.

[19]. Tanaka, K., and Iwasaki, H,Extraction of lexical translations from non-aligned corpora In Proceedings of the 16th International Conference on Computational Linguistics, 580, 1996.

[20]. Karaivanov,Svetoslav,Veselin and Martin, Phrase-Based Statistical translation Of Programming Languages in proceedings of 2014 ACM international Symposium on New ideas New Paradigms and Reflections on Programming And Software,2014.

[21]. acl.eldoc.ub.rug.nl