**HINDI-ENGLISH MACHINE TRANSLATION SYSTEM USING BAYESIAN WORD ALIGNMENT APPROACH WITH GIBBS SAMPLING**

Atul Rauthan a, Vinod Kumar Mishra b

Department of Computer Science & Engineering,

B.T. Kumaon Institute of Technology, Dwarahat, Uttarakhand-263653, India

Email:-[atulrauthan17@gmail.com](mailto:atulrauthan17@gmail.com)a, [vkmishra2005@gmail.com](mailto:vkmishra2005@gmail.com) b

**ABSTRACT**

**The Hindi-English machine translation system is used to translate Hindi to English and vice-versa. Hindi is a SubjectObjectVerb order type language while English is SubjectVerbObject order type language. Different word order causes a fair amount of complexity in the machine translation of Hindi-English. The Hindi-English machine translation systems mostly use the expectation maximization (EM) algorithm for the word alignment but the BiLingual Evaluation Understudy (BLEU) scores is not significant in this method . Therefore in order to improve the BLUE score i.e. to increase the accuracy of Hindi-English machine translation system, we propose a Bayesian word alignment approach using the Gibbs Sampling (GS) algorithm. This approach improves the quality of the translation in Hindi-English machine translation system. Gibbs sampling is use to sample the word alignments, obtain by processing the parallel corpora. Word alignments are compare against EM and Variational Bayes (VB) alignments in terms of the end-to-end translations. Performance is compare on Hindi-English translation bi-directionally. The quality of the translations is measure using the BiLingual Evaluation Understudy (BLEU) scores. The results indicate that the Bayesian approach of word alignment with Gibbs sampling outperforms in comparison of both EM and VB in terms of BLEU scores.**

**Keywords: Hindi-English Machine Translation; Gibbs Sampling; Bayesian alignment; Expectation Maximization; Variational Bayes.**

**INTRODUCTION**

Machine Translation (MT) is a sub-field of computational linguistics that investigates the use of computing resources to translate text or speech from one natural language to another.

At its basic level, MT performs simple substitution of words in one natural language for words in another language. Current machine translation software often allows for customization by domain or profession (such as weather reports, news translations), improving output by limiting the scope of allowable substitutions in the system. This technique is particularly effective in domains where formal or formulaic language is used. It follows that machine translation of government and legal documents more readily produces usable output than conversation or less standardized text.

Word alignment is a very crucial and important step in the machine translation system. Main goal of word alignment is to identify the mapping between the source and the target words in the parallel sentences, when sentence aligned parallel corpus is given. The word alignment information is not usually available during the corpus generation. The task of word alignment is an unsupervised learning problem.

The end to end translation quality of a machine translation system is majorly dependent on the word alignments inferred in the word alignment phase. Most used machine translation approach is statistical machine translation (SMT) [1] in which statistical techniques are used to find the word alignments. SMT uses expectation maximization [EM] [2] algorithm to align the words of given parallel corpus.

In this paper we describe a system which uses the Bayesian word alignment using gibbs sampling algorithm to improve the quality of the translation. Our system is motivated by the desire to develop an improved alignment system for the English and Hindi language pair for the language translators. With improved alignments, the quality of end to end translations will improved. We compare the quality of end to end translation using the bilingual evaluation understudy (BLEU) scores.

The rest of the paper is organize as follows: In the next section we describe the literature review, section 3 gives the detail of background and experimental setup while section 4 describes the detail of the outcome of simulation results and article ends with summary and conclusions of the article with future works.

**LITERATURE REVIEW**

The first idea of the statistical machine translation was introduced by Warren Weaver in year 1949. He introduced the idea of applying Claude Shannon's information theory[3] in the SMT. Statistical machine translation was re-introduced in 1993 by researchers Thomas J. Watson at IBM's Research Centre [4]. Nowadays it is by far the most widely studied machine translation method.The basic idea behind the Statistical Machine translation comes from information theory. A document is translated according to the probability distribution which shows that a string  in the target language (for example, English) is translation of a string in the source language (for example, Hindi). The problem of modeling the probability distribution  is crucial in SMT. One approach which lends itself well to computer implementation is to apply Bayes Theorem, that is, where the translation model  is the probability that the source string is the translation of the target string and the language model  is the probability of getting that target language string after the translation. Finding the best translation is done by picking up the one translation that gives the highest probability:

 .

In the above process there is a crucial step named the word alignment process. The word alignment is done using the IBM models of the word alignments. IBM model-1[4] gives the alignment algorithm for aligning the words when given the parallel corpora. Hindi-English machine translation has been a center of research for the computer linguistics in Indian Sub continent. Many systems has been developed as Anglabharati (and Anubharati), MaTra, UNL-based MT between the English, Hindi and Marathi language and Sampark system. All theseprojects use the EM algorithm implementation of the IBM models for the word alignment.

According to Xu et al., a Dirichlet Process prior is placed on IBM Model 1 word translation probabilities in word alignment phase [5]. In Nguyen et al., a Pitman-Yor Process prior is placed on word translation probabilities in a proposed bag-of-words translation model that is similar to IBM Model 1 [6]. Both studies utilized Gibbs sampling for inference. However, alignment distributions were not sampled from the true posteriors but instead were updated either by running GIZA++ (Xu et al.) [5] or by using a “local-best” maximization search Nguyen et al. [6]. On the other hand, a sparse Dirichlet prior on the multinomial parameters was used in (Chung et al.) [7] to prevent over fitting.

Bayesian word alignment with Dirichlet priors was also investigated in a recent study using Variational Bayes (VB) (Riley et al.) [8].VB is a Bayesian inference method which is sometimes preferred over Gibbs sampling due to its relatively lower computational cost and scalability. However, VB inference approximates the model by assuming independence between the hidden variables and the parameters. Vogel et al. described a new model for word alignment in statistical translation. In their model the alignment probabilities were dependent on the differences in the alignment positions rather than on the absolute positions [9].

**BACKGROUNDAND EXPERIMENTAL SETUP**

The machine translation system has a very important phase as the word alignment phase which takes parallel corpus as the input and outputs the alignments and word translation probabilities. The alignment algorithms use some models which provide the procedures on how the words should be aligned in the machine translation step. As such is the IBM model of word alignment.

**IBM model of word alignment**

The IBM’s word alignment model [4] is used to align the words from the sentences aligned parallel corpus. Given a parallel corpus(E,F) of S sentence pairs, let e(f) be the sth sentence pair in E(F) and ei(fj) denote the ith(jth)word among a total of I(J)words ine(f). Let there be a “NULL” word e0 to account for any unaligned words in f. let VE and VF denote the size of the respective vocabularies. With each fj we associate a hidden alignment variable aj whose value ranges over [0, I]. The set of alignments for a sentence (corpus) is denoted by a(A). The model parameters consists of a VE ×VF table T of the word translation probabilities such that te,f=P(f|e). Since f is conditional on e, we refer to e(e)as the “source” word (sentence) and f(f) as the “target” word(sentence).

The conditional distribution of the IBM model 1 is:

aj|e ̴ Uniform (aj;I+1)

P(F,A|E;T)= (1)

= (2)

The unknowns A and T are estimated using the Expectation Maximization algorithm. It finds the value of T that maximizes the likelihood of the observed variables E and F according to the model.

P(F,A|E,T)= (3)

= (4)

The count variable ne,f,s denotes the number of times the source word type e is aligned to the target word type f in the sentence pair s, and in (4) Ne,f=.

**Gibbs Sampling**

Suppose we want to obtain k samples of from a joint distribution . We denote ith sample by . Here in the machine translation we use the Gibbs sampling to sample the alignments distributions P(A|E,F;θ). For a set of random variables z={zj}, a Gibbs sampler iteratively updates the variable zj on at a time by sampling the value from the distribution P(zj|z⌝j), where the superscript ⌝j denotes the exclusion of the variable being sampled currently.

P(F,A|E;θ)= (5)

The Gibbs sampling formula [10] is as:

P(aj =i|E,F,A⌝j;θ)∝ (6)

Here denotes the number of the times the source word type ei is aligned to the target word type fj in A, not taking the current alignment link between fj and in account.

The algorithm for Gibbs sampling is:

Input E,F; Output K samples of alignment A

1. Initialize A
2. For k=1 to K do
3. For each sentence pair s in (E,F) do
4. For j=1 to J do
5. For i=0 to I do
6. Calculate P(aj=i|………)according to (6)
7. Sample a new value for aj.

We evaluated the performance of the Bayesian word alignment via bi-directional experiment from Hindi to English and English to Hindi. Firstly we setup the system for statistical machine translation using the Moses [11], GIZA++ [12] and SRILM [13] using the expectation maximization algorithm. Language models use are 4-gram with the modified Kneser-Ney smoothing [14].

For Hindi-English experiments, we used the Hindi-English without Emille parallel corpus from the lindat repository [15]. We used the news-commentary-v8.hi-en for the language model training and training the translation system. We used news-test2008 for the tuning of the translation system. We finally used the newstest2011 corpus to check the translation system and comparing the quality of the translation.

|  |  |
| --- | --- |
| Corpus | Number |
| Training Set:  Sentences  Tokens  Tuning Set:  Sentences  Tokens  Test Set:  Sentences  Tokens | 700  14433  2736  50015  755  15499 |

We firstly preprocessed the corpuses. Firstly the corpus was tokenised and the tokenized corpus was then true cased. The true cased corpus was then cleaned. The long sentences which were having words more than 60 were removed from the corpuses. After the preprocessing the further processing was done with the corpuses. For each language pair we obtained maximum-likelihood word alignments using the Expectation maximization implementation of GIZA++ and Bayesian alignments using the Gibbs sampling. After alignments were obtained in both the translation directions, standard phrase-based SMT systems were trained in both directions using Moses, SRILM, and MERT [16] tools. Translations were evaluated using the single-reference BLEU [17] metric. For Gibbs Sampling we firstly aligned words in the parallel corpus using the Expectation Maximization and then applied the Gibbs Sampling on that aligned words.

We used the publicly available software [18] for the Variational Bayes implementation of the word alignment step. Then we compared the results of all the three different alignment algorithms.

We compared the quality of the translation on the basis of the BLEU scores calculated on the end to end translations in the both Hindi to English and English to Hindi Statistical Machine Translation Systems.

**RESULTS**

We experimented with EM, VB and GS inferred alignment algorithm with the same corpus and input texts.

Table 1 shows the BLEU scores of SMT system trained with the individual EM, VB and GS inferred alignments for the Hindi to English machine translations. The GS alignments lead to higher BLEU scores on averagethan using the EM alignments directly. The VB alignments also lead to the higher BLEU scores than EM alignments. But the best results are provided using the GS alignments.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Algo. | Overall BLEU | Unigram | Bigram | Trigram | Fourgram |
| EM | 2.95 | 24.1 | 3.3 | 1.3 | 0.8 |
| VB | 3.15 | 24.3 | 3.6 | 1.3 | 0.8 |
| GS | 3.77 | 24.5 | 3.9 | 1.7 | 1.2 |

**Table: - 1 BLEU scores of SMT for Hindi to English translation using different alignment algorithms**.

Table 2 shows the BLEU scores of SMT system trained with individual EM, VB and GS inferred alignments for the English to Hindi translations. The GS alignments lead to higher BLEU scores on average than using the EM alignments directly. The VB alignments also lead to the higher BLEU scores than EM alignments. But the best results are provided using the GS alignments.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Algo. | Overall BLEU | Unigram | Bigram | Trigram | Fourgram |
| EM | 1.03 | 13.8 | 1.6 | 0.2 | 0.1 |
| VB | 1.12 | 16.4 | 2.0 | 0.4 | 0.1 |
| GS | 1.61 | 19.4 | 2.3 | 0.6 | 0.6 |

**Table:-2 BLEU scores of SMT for Hindi to English translation using different alignment algorithms.**

Fig. 1 shows the comparison of the BLEU scores of SMT systems trained on the individual EM, VB and GS alignments. In all the cases, the GS alignments which are initialized with the alignments from EM leads to the higher BLEU scores on average than using the EM alignments directly on the overall corpus. But for the Unigram BLEU score on Hindi to English translation, it decreases slightly in the case of the Gibbs Sampling. It is because of the rare word fertility and the one to many translation of the Hindi to English translations.

**Fig:-1 BLEU scores for the Hindi to English Statistical Machine Translation system.**

The translation system for English to Hindi machine translation using the GS provides the best results in terms of end to end translations. Fig 2 shows the comparison of BLEU scores of the English to Hindi SMT.

**Fig:-2 BLEU scores for the English to Hindi Statistical Machine Translation system.**

The translation performance of the individual VB alignments, compared to EM, VB achieves higher BLEU scores in HE and EH. GS outperforms VB in all cases. We found that the Variational Bayesian gives the improvement on the unaligned words in Hindi-English alignment on the EM algorithm.GS outperforms both the EM and VB in both the English to Hindi and the Hindi to English machine translation system.

**CONCLUSION**

In this paper, we developed a machine translation system using Gibbs sampling-based word alignment and compared the results with the EM estimation in terms of the end to end translation BLEU scores in the Hindi-English Statistical Machine Translation systems. The result shows significant improvements in the cases of the sparse data as in the cases of the smaller corpora and more morphological complexity. This proposed method successfully overcomes the Rare Word Fertility Problem in the EM–estimated word alignments. The results also indicate that the Gibbs sampling perform better than Variational Bayes Inference in both Hindi to English and English to Hindi SMT systems. In future we plan to work on further improvement of the alignments using the part of speech information for the Hindi and the English texts and to improve the quality of the translations we will use the morphologically processed corpuses. We will also work on the larger corpus to see that whether VB inference and Gibbs sampling gives improved results or not. It also includes the estimation of the hyper parameters from data sources and utilization of the proposed algorithms in either initialization or inference of the advanced alignment models.

**REFERENCES**

**[1]** P. Koehn*,* “Statistical Machine Translation”, Cambridge, U.K.: Cambridge Univ. Press, 2010.

**[2]** A. Dempster, N. Laird and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm”,J. R. Statist.Soc., Series B, vol. 39, no. 1, pp. 1–38, 1977.

**[3]** C. E. Shannon, and W. Weaver, “The Mathematical Theory of Communication”, Urbana: The University of Illinois Press. 1949.

**[4]** P. F. Brown, V. J. D. Pietra, S. A. D.Pietra, and R. L. Mercer, “The mathematics of statistical machine translation: Parameter estimation,” Comput. Linguist, vol. 19, no. 2, pp. 263–311, 1993.

**[5]** J. Xu, J. Gao, K. Toutanova and H. Ney, “Bayesian semi-supervised Chinese word segmentation for statistical machine translation”, in Proc. COLING, Manchester, U.K., pp. 1017–1024, Aug. 2008.

**[6]** T. Nguyen, S. Vogel and N. A. Smith, “Nonparametric word segmentation for machine translation”, in Proceeding of COLING, pp. 815–823, 2010.

**[7]** T. Chung and D. Gildea, “Unsupervised tokenization for machine translation”, in Proc. EMNLP, Singapore, pp. 718–726, Aug. 2009.

**[8]** D. Riley and D. Gildea, “Improving the IBM alignment models using variational Bayes”, in Proceedings of ACL: Short Papers, pp. 306–310, 2012.

**[9]** S. Vogel, H. Ney and C. Tillmann, "HMM-based word alignment in statistical translation", Proceedings of the 16th conference on Computational linguistics, Vol. 2, Association for Computational Linguistics, pp. 836-841, 1996.

**[10]** C. Mermer and M. Saraçlar,“Improving Statistical Machine Translation Using Bayesian Word Alignment and Gibbs Sampling”, ieee transactions on audio, speech, and language processing, vol. 21, no. 5, pp. 1090-1101, may 2013

**[11]** P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, “Moses: Open source toolkit for statistical machine translation,” in Proc. ACL: Demo and Poster Sessions, Prague, Czech Republic, pp. 177–180, Jun. 2007.

**[12]** F. J. Och and H. Ney, “A systematic comparison of various statistical alignment models”, Comput. Linguist., vol. 29, no. 1, pp. 19–51, 2003.

**[13]** A. Stolcke, “SRILM—An extensible language modeling toolkit,” in Proc. ICSLP, vol.3, 2002.

**[14]** S. F. Chen and J. Goodman, “An empirical study of smoothing techniques for language modeling”, Comput. Speech Lang., vol. 13, pp. 359–394, 1999.

**[15]**[Online]Available: <https://lindat.mff.cuni.cz/repository/xmlui/handle/11858/00-097C-0000-0001-BD17-1>

**[16]** J. H. Clark, C. Dyer, A. Lavie and N. A. Smith, “Better hypothesis testing for statistical machine translation: Controlling for optimizer instability”, in Proc. ACL: HLT, Portland, USA, pp. 176–181, 2011.

**[17]** K. Papineni, S. Roukos, T. Ward and W.J. Zhu, “BLEU: Amethod for automatic evaluation of machine translation,” in Proc. ACL, Philadelphia, PA, USA, pp. 311–318, Jul. 2002.

**[18]** D. Riley and D. Gildea, “Improving the performance of GIZA++ using variational Bayes”, Univ. of Rochester, Comput. Sci. Dept., Rochester,NY, USA, Tech. Rep. 963, Dec. 2010.