Volume 2, Issue 8, August 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Perspective to Enhance the Quality of Translation through Multi-Engine Machine Translation (MEMT) System

Pankaj Goswami*
AMITY University, India

Sanjay Dwivedi
BBA University Lucknow India

Abstract— Multi Engine Machine Translation (MEMT) system is an approach where more than one suitable translation engines are synthetically combined for enhancing the quality of machine translation. This system is tested on many language pairs in world wide and produced significantly encouraging results. Our objective is to explore the suitability of MEMT in high quality translation for English to Hindi.

Keywords— Multi Engine Machine Translation (MEMT), Machine Translation (MT), Rule Based Machine Translation (RBMT), Statistical Machine Translation (SMT), Example Based Machine Translation (EBMT).

I. INTRODUCTION

The penetration of machine translation has reached into most localize form and the requirement of correct translation would be the basic requirement of any language pair system. Currently various research based machine translation systems are available and they are individually giving good results like Statistical Machine Translation (SMT), Rule Based Machine Translation (RBMT) and Example Based Machine Translation (EBMT) but they are not reaching up to the optimal quality translation as required.

Our research based on the initial concept that "three heads are better than one" [12]. In this hypothesis, multiple translation engines works on same input text and they individually translate and produce their respective output text after that a strong combining algorithm produce the final and optimal translation.

In this study we are trying to find out the gradual development in MEMT and explorer the suitability for English to Hindi translation.

II. RELATED STUDIES IN MEMT (YEAR WISE)

1994: An initial study done by R. Frederking and Sergei Nirenburg that "Three heads are better than one" and applied it on Spanish to English language pair translation. As multiple heads they incorporated Knowledge Base MT, Example Based MT, Lexical transfer MT to enhance the translation quality. A chart-walk algorithm helps to produce single, best, non-overlapping, contiguous combination from the available component translations. *The concluding remarks that a multi-engine system depends on the basic quality of each particular engine* [12].

Further in the same year study published by Robert Frederking and Sergei Nirenburg et al. on extended idea of "Integrating Translations from Multiple Sources With in the Pangloss Mark III Machine Translation System" and applied it on Japanese to English translation. This system is based on Multi Engines like Knowledge Base MT, Example Based MT, Lexical transfer MT. The component translations reformed by tested algorithm Chart-walk, to produce a single, best, non-overlapping, contiguous combination. As a conclusion multi-engine system depends on the basic quality of each particular engine. They expect the individual engines like KBMT and EBMT to perform best. They raised very important point that *the multi-engine environment will be improved, when larger static knowledge sources are added and the scoring mechanism is further adjusted* [13].

Furthermore in same year an extensive study "Toward Multi-Engine Machine Translation" performed by Sergei Nirenburg Robert Frederking on Spanish to English language pair. The same combination engines were used Knowledge Base MT, Example Based MT, Lexical transfer MT and the same Chart-walk algorithm to produce a single, best, non-overlapping, contiguous combination of the available component translation. A multi-engine system depends on the basic quality of each particular engine. They stressed on the performance of individual engines (especially, KBMT and EBMT). They expressed their views about *the essence of an automatic testing procedure that would assessed the utility of the multi-engine system relative to the engines taken separately* [14].

2001: An extensive study done by Chris Callison-Burch, Raymond S. Flournoy- "A program for automatically selecting the best output from multiple machine translation engines", on language pairs Japanese to English and French to English. The specific engine names are not mentioned. The best translation obtained through one crucial assumption: *that the most fluent output corresponds to the best translation*. Through test it is been observed that the assumption performed upto 19% better than the baseline metric in machine translation [5].

The contributory concept for enhancing translation quality through MEMT given by Fuji Ren, Hongchi Shi Shingo Kuroiwa in the study "A New Machine Translation Approach Using Multiple Translation Engines and Sentence Partitioning" with the test language Chinese to Japanese and multi engines: Rule based Translation Engine, Example Based Translation Engine, Family Model based translation Engine, Super function based translation Engine. The Translation quality can be improved by incorporating a sentence partitioning and coordination mechanism. Partitioning complex sentences plays an important role in MT, especially for Chinese language. In this study researchers found that *the main reasons for low translation quality is the difficulty in finding the correct translation from so many candidates* [8].

The result shows that the correct translation rate without multiple MT engines is 77.4%, while the rate with multiple MT engines is 84.3%. Another experiment shows the correct translation rate without multiple MT engines is 45.6% while the rate with multiple MT engines is 74.2%.

They also tested translation speed and found without multiple MT engines takes 54 seconds to translate (15171 Japanese words), while the multiple MT engines takes 98 seconds. *The multiple MT engine approach slows down the translation speed in less than two times but it improves the translation quality* [8].

2002: Subsequent study performed by Yasuhiro Akiba, Taro Watanabe and Eiichiro Sumita "Using Language and Translation Models to Select the Best among Outputs from Multiple MT systems" applied on Japanese and English (J-E and E-J) language pair. Multiple systems TDMT, D3, SMT were used for (J-E) pair and TDMT, HPAT, SMT for (E-J) pair of language. This study addressed the challenging problem of automatic selection of the best among outputs from multiple MT systems to improve translation quality. The performance of proposed methods is much better than the existing methods and improvement of 2 to 6% is recorded [18].

2004: Tadashi Nomoto, expressed "Multi-Engine Machine Translation with Voted Language Model", using English to Japanese language pair. In this study two confidence models: Fluency based Model (FLM) and Alignment based Model (ALM) been used. The choice of language model could affect performance of MEMT. One possible approach would be the use of dynamic language model by re-training itself on data sampled from the web [15].

2005: A different angle in the same context given by Shyamsundar Jayaraman and Alon Lavie "Multi-Engine Machine Translation Guided by Explicit word Matching" and applied on Chinese to English language pair. To achieve better quality translation three online machine translation systems were combined Systran, Netat, Wordlingo. A new approach for synthetically combining the output of several different Machine Translation (MT) engines operating on the same input. The goal is to produce a synthetic combination that surpasses all of the original systems in translation quality.

Through experiments the new multi-engine combination system achieves an improvement of about 6% over the best original system. MEMT decoder produces hypotheses that are even far superior in translation quality, but the current scoring algorithm is not yet capable of selecting the best generated hypothesis [9].

In the same year a study published by M. van Zaanen and H. Somers "DEMOCRAT: deciding between multiple outputs created by automatic translation" tested on French to English language pair. Online multiple translation systems combined to improve quality tranlstion: Babelfish, Freetranslation, Systran, TranslateRU (ProMT) and Worldlingo. This system tested on the output of free on-line MT systems. *The result proves that no individual MT system is the best. While DEMOCRAT (DEcides between Multiple Outputs CReated by Automatic Translation) may not always beats the best individual MT system* [11].

2006: In same line of research the study done by Bart Mellebeek, Karolina Owczarzak et al. "Multi-Engine Machine Translation by Recursive Sentence Decomposition" using English to Spanish language pair. The multi engine like LogoMedia, Systran, SDL applied. Their approach is based on a recursive decomposition of the input sentence into smaller chunks which are more likely to be correctly translated than the longer input sentence [4].

2007: A remarkable study done by Antti-Veikko I. Rosti and Necip Fazil et al. "Combining Outputs from Multiple Machine Translation Systems" tested on Arabic to English and Chinese to English. Six systems (Unnamed, A,B,C,D,E,F) trained on all available data. Three system were phrase-based (A,C,E), two hierarchical (B and D) and one syntax-based (F). The outputs from six very different MT systems, tuned for two different evaluation metrics, may be combined to yield better outputs in terms of different evaluation metrics [3].

In the same year a study published by Yu Chen, Andreas Eisele et al. "Multi-Engine Machine Translation with an Open-Source Decoder for Statistical Machine Translation" and tested on German to English language pair. Multi Engines used in this research are Statistical Machine Translation (SMT) with Rule Based Machine Translation (RBMT). The input text and the output text of the MT systems was aligned by means of GIZA++, a tool with which statistical models for alignment of parallel texts can be trained [19].

2008: A view of combining "Hybrid Machine Translation Architectures within and beyond the EuroMatrix project" given by Andreas Eisele, Christian Federmann et al. applied on English paired with Spanish and German in both directions. Multi engines are Rule Based Machine Translation (RBMT), Statistical Machine Translation (SMT) used. A fluency model can be integrated into RBMT based architecture via post-editing [2].

In the same year Evgeny Matusov and Gregor Leusch et al presented "System Combination for Machine Translation of Spoken and Written Language" tested on Spanish to English and from English to Spanish. In this study six state-of-the-art statistical phrase-based translation systems**:** *Phrase-Based Model, Word-Based Lexicon Model, Word and Phrase Penalty, RWTH:* The SMT system, *UKA:* The UKA phrase-based SMT system, UPC: The SMT system different from the other partner systems. A trigram language model trained on the six system translations for each of the 1130 evaluation data sentences (on 6780 sentences). Well-established automatic evaluation measures like the BLEU score, word error rate (WER), position-independent word error rate (PER), and the NIST score were calculated to assess the translation quality [7].

2009: In a view of system combination Xiaodong HE and Mei Yang et al. presented “Improved Monolingual Hypothesis Alignment for Machine Translation System Combination” for Chinese to English language pair. As a combination eight systems were combined [17].

Yu Chen and Michael Jellinghaus et al. presented their study in same year “Combining Multi-Engine Translations with Moses” with six pairs of translation between four languages German, French, Spanish and English. They collect translations from all available systems and pair them with the corresponding input text, thus forming a medium-sized “hypothesis” corpus. This system starts processing this corpus with a standard phrase-based SMT setup, using the Moses toolkit. In this work Moses toolkit used to combine translations from multiple engines in a simple way. The combination performs better than the best system in the half of the six translations [20].

2010: WenpengLu and Ruojuan Xue presented their view on “Comparative Study on Multi-systems Combination in Machine Translation” tested on German language. The RBMT and SMT system are the component engines. They compared three methods of combining MT system- serial, parallel and hybrid. They would select a RBMT system to dominate the syntactic structure and import the language model of SMT system to get the optimal combination and improve the overall translation performance [16].

In the same year Kenneth Heafield and Alon Lavie presented their study “CMU Multi-Engine Machine Translation for WMT 2010” with participation of nine pairs of English, Czech, French, German and Spanish language. In this experiment they include 5 to 17 numbers of systems to combine and produce highest scoring through parameters like BLEU, TER, METE. The candidate combinations are scored by their length, agreement with the underlying system and language models-

Czech: CzEng,
English: Gigaword-IV(Fr-En and CzEng),
French: Gigaword-II(Fr-En),
Spanish: Gigaword-II

The study shows that performance of the system will improve for multiple languages and the improvement in BLEU over the best system depends on the language pair it ranges from 0.89% to 5.57% with mean 2.37% [10].

III. MULTI-ENGINES VS LANGUAGE PAIRS

We found that in most of the MT system, English language would be preferred as one of the pairing language. It has been observed that researchers are very much inclined to achieve good quality translation in their native language from the foreign languages. On the basis of many research papers we would be able to say that multi engine machine translation adopted by many native languages as shown in Table I.

In this journey of review we found that quality translation work though MEMT not yet started for Hindi native language. The promising results of MEMT motivated us to explore the scope of this approach in Hindi MT system.

TABLE I.

PERIODICAL SUMMARY OF MEMT (1994-2010)

Year	Language Pair	Multi Engines Used
1994	Spanish to English	KBMT, EBMT, Lexical transfer MT
	Japanese to English	KBMT, EBMT, Lexical transfer MT
2001	Chinese to Japanese	RBMT, EBMT, Family Model based Engine, Super function based Engine.
2002	Japanese to English and English to Japanese	(J-E): TDMT, D3, SMT. (E-J): TDMT, HPAT, SMT
2004	English to Japanese	Fluency based Model (FLM) and Alignment based Model (ALM)
2005	French to English	Babelfish, Freetranslation, Systran, TranslateRU (ProMT), Worldlingo.
	Chinese to English	Systran, Netat, Wordlingo.
2006	English to Spanish	LogoMedia, Systran, SDL.
2007	Arabic to English and Chinese to English	Six systems (Unnamed, A,B,C,D,E,F). Three system were phrase-based (A,C,E), two hierarchical (B and D) and one syntax-based (F).
	German to English	Statistical Machine Translation (SMT) with Rule Based Machine Translation (RBMT)
2008	English to Spanish English to German	Rule Based Machine Translation (RBMT), Statistical Machine Translation (SMT)

	Spanish to English and English to Spanish	Six statistical phrase-based translation systems**:** Phrase-Based Model, Word-Based Lexicon Model, Word and Phrase Penalty, RWTH: The SMT system, UKA: phrase-based SMT system, UPC: The SMT system.
2009	Chinese to English	Tree-to-String System, Phrase-Based System, Syntactic Source Reordering System, Syntax-Based Pre-Ordering System, Hierarchical Phrase-Based System, Lexicalized Reordering System, Two-Pass Phrase-Based System, Hierarchical System.
	(German, French, Spanish) to English	Standard phrase based SMT framework and RBMT system.
2010	English to (Czech, French, German, Spanish).	Included 5 to 17 numbers of systems to combine and produce highest scoring through parameters like BLEU, TER, METE
	German	Rule Based Machine Translation (RBMT), Statistical Machine Translation (SMT)

IV. COMPILATION OF VARIOUS TEST RESULTS

In this section we are trying to sum up the results of various researches done in more than one decade. In the initial stage of MEMT Sergei Nirenburg found that the system depends on the basic quality of each particular engine. It is useful to have an automatic testing procedure that would assess the utility of the multi-engine system relative to the engines taken separately. The multi-engine environment will improve, as larger static knowledge sources being added and the scoring mechanism may be further adjusted [12]. The most fluent output corresponds to the best translation, this assumption performed up to 19% better results than the baseline metric [5].

In a different experiment, where translation speed found with out multiple MT engines takes 54 seconds to translate (15171 Japanese words), while the multiple MT engines takes 98 seconds. *The multiple MT engine approach slows down the translation speed by less than two times but it improves the translation quality* [8].

The performance of the proposed methods (Kruskal-Wallis test and conditional probability) is much better than that of the existing methods and improvement recorded 2 to 6% in performance [18]. The choice of language model could affect the performance of MEMT [15]. Jayaraman and Alon Lavie have done an experiment on multi engine combination system. This achieves an improvement of about 6% over the best original system [9].

The automatic system DEMOCRAT (DEcides between Multiple Outputs CReated by Automatic Translation) tested on the output of free on-line MT systems. The result proves that no individual MT system is the best [11]. MEMT results are showing statistical significant of relative improvements up to 9% in BLEU score [4].

An experiment shows that out of sentence-level and Phrase-level combinations the word-level combination method based on consensus network decoding seems to be very robust and yield good gains over the best single system [3]. According to Yu Chen, the multi-engine MT driven by a SMT decoder placed a strong emphasis on the statistical models [19].

The rule-based MT engines are used to enrich the lexical resources available to the SMT decoder. In other case, parts of the SMT infrastructure are used, together with linguistic processing and manual validation, to extend the lexicon of a rule-based MT engine. Both approaches have been implemented and show promising improvements to MT quality [2].

In a different work with Moses toolkit used to combine translations from multiple engines in a simple way. The combination performs better than the best system [20]. To get the optimal combination and improvement in the overall translation performance it should select a RBMT system to dominate the syntactic structure and import the language model of SMT system [16]. The MEMT system will improve BLEU score over the best system depending on the language pair and its ranges from 0.89% to 5.57% with mean of 2.37% [10].

V. METHODOLOGIES ADOPTED BY DIFFERENT RESEARCHES

On the MEMT approach various researches have already been done. We are trying to present a glimpse of methodologies adopted by different researches. Table-II shows the summary of research methodologies adopted by various research studies done so far in this domain.

TABLE II.

MAJOR METHODOLOGIES ADOPTED IN MEMT

Researchers	Major Methodologies	Concluding Remarks
Kenneth Heafield, Alon Lavie (2010)	Included 5 to 17 numbers of systems to combine and produce highest scoring through parameters like BLEU, TER, METEOR.	The improvement ranges from 0.89% to 5.57% with mean 2.37% in BLEU score over the best system depending on the language pair.
WenpengLu, Ruojuan Xue (2010)	Compared three methods of combining MT system: serial, parallel and hybrid.	Hybrid combination produced better results, especially SMT to RBMT.
Yu Chen, Michael Jellinghaus et al. (2009)	Moses toolkit used to combine translations from multiple engines.	Promising improvements shown in terms of BLEU scores increase up to 4.25 points.

Xiaodong HE, Mei Yang et al. (2009)	Indirect Hidden Markov Model (IHMM) based system combination presented and confusion network incorporated.	IHMM based hypothesis alignment method provided superior results as compare to the TER (translation error rate) based method.
Evgeny Matusov, Gregor Leusch et al, (2008)	The consensus translation is computed by weighted majority voting on a confusion network. A trigram language model trained on the six system translations.	Automatic evaluation measures like the BLEU score, Word Error Rate (WER), Position independent Word Error Rate (PER), and the NIST score used.
Andreas Eisele, Christian Federmann et al. (2008)	Rule-based MT engines are used to enrich the lexical resources available to the SMT decoder. SMT used together with linguistic processing and manual validation, to extend the lexicon of a rule-based MT engine.	An architecture required which has simultaneous access of all knowledge, which is beyond the presented architecture.
Yu Chen, Andreas Eisele et al. (2007)	The input text and the output text of the MT systems aligned by GIZA++, a tool with statistical models.	The described architecture placed a strong emphasis on the statistical models.
Antti-Veikko I. Rosti, Necip Fazil Ayan et al. (2007)	Presented three combination methods. Sentence level combination, Phrase level combination, word level combination.	The word-level combination method based on consensus network decoding seems to be very robust and yield good results.
Bart Mellebeek, Karolina Owczarzak et al. (2006)	Recursive decomposition of the input sentence into smaller chunks. A selection procedure based on majority voting. The best chunk translations are then recomposed into target language sentences.	Results are statistically significant, relative improvements of up to 9% BLEU score.
Shyamsudar Jayaraman and Alon Lavie (2005)	An explicit *Word Matcher*. A decoding algorithm. Confidence estimates for the various engines and Trigram language model used.	This combination improves 6% over the best original system. Potential to work on scoring mechanism with decoder and improvement required in word matcher.
M. van Zaanen and H. Somers (2005)	Aligns all the input sentences. Analyses the alignments and builds a graph. Walk through the graph and generates the output sentence.	The system tested on the output of free on-line MT systems. The result proves that no individual MT system is the best.
Tadashi Nomoto (2004)	Voting by Majority (V by M) to choose among Language Models.	The choice of language model could affect the performance of MEMT.
Yasuhiro Akiba, Taro Watanabe et al. (2002)	Kruskal-Wallis test and conditional probability used.	An improvement recorded 2 to 6% in performance.
Fuji Ren, Hongchi Shi Shingo Kuroiwa (2001)	Sentence partitioning and coordination mechanism.	The multiple MT engine approach slows down the translation speed in less than two times but it improves the translation quality.
Callison-Burch C., Raymond S. Flournoy, (2001)	Most fluent output corresponds to the best translation.	The assumption performed up to 19% better than the baseline metric.
R. Frederking, Sergei Nirenburg et al. (1994)	Chart-walk algorithm	A multi-engine system depends on the basic quality of each particular engine with larger static knowledge as well as good best scoring mechanism.

VI. CONCLUSION

In this review study we found that most of the native language researchers are working for better Machine Translation quality and they are eager to evolve different possible solution. MEMT would definitely be a perfect approach for quality machine translation.

Various researches are showing encouraging results that is why we are hopeful to achieve good results of Multi-Engine Machine Translation in Hindi native language.

It is being observed that as multi-engine components, SMT was the first appropriate choice and others like RBMT, EBMT were the next. These set of engines would be preferable choice for our further extension of research work towards Hindi translation.

REFERENCES

[1] Andreas Eisele, “First Steps towards Multi- EngineMachine Translation”, *in Proc. of the ACL Workshop on Building and Using Parallel Texts*, 2005, p. 155–158.

[2] Andreas Eisele et al., “Hybrid Machine Translation Architectures within and beyond the EuroMatrix project”, *in Proc. of the 12th EAMT conference*, 2008, p. 27-34.

[3] A.-V. I. Rosti, N. F. Ayan, B. Xiang, S. Matsoukas, R. Schwartz, and B. J. Dorr, “Combining Outputs from Multiple Machine TranslationSystems”, *in Proc. of NAACL HLT*, 2007, p. 228-235.

[4] Bart Mellebeek et al., “Multi-Engine Machine Translation by Recursive Sentence Decomposition”, *in Proc. of the 7th Conference of the Association for Machine Translation in the Americas*, 2006, p. 110-118.

[5] Callison-Burch, C., Flournoy, R.S, “A program for automatically selecting the best output from multiple machine translation engines”, *in Proc. of MT Summit VIII*, 2001, p. 63-66.

[6] E. Matusov, N. Ueffing, and H. Ney, "Computing Consensus Translation from Multiple Machine Translation Systems Using Enhanced Hypotheses Alignment", *in Proc. of EACL*, 2006, p. 33-40.

[7] Evgeny Matusov and Gregor Leusch et al, "System Combination for Machine Translation of Spoken and Written Language*", IEEE transactions on audio, speech, and language processing*, vol. 16, no. 7, pp.1222-1237, 2008.

[8] Fuji Ren, Hongchi Shi, Shingo Kuroiwa, "A New Machine Translation Approach Using Multiple Translation Engines and Sentence Partitioning", *in proc. of IEEE International Conference on systems*, man and cybernetics, 2001, p.1699-1704.

[9] Jayaraman S. et al., "Multi-Engine Machine Translation Guided by Explicit Word Matching", *in proc. of EAMT-2005 Conference*, 2005.

[10] Kenneth Heafield, Alon Lavie, "CMU Multi-Engine Machine Translation for WMT 2010", *in Proc. of Workshop on Statistical Machine Translation*, 2010, p. 301–306.

[11] M. van Zaanen and H. Somers, "DEMOCRAT: deciding between multiple outputs created by automatic translation". *In Proc. of the 10th Machine Translation Summit*, 2005, p. 173–180.

[12] R. Frederking, S. Nirenburg, "Three heads are better than one", *in Proc. of the fourth conference on Applied natural language processing*, Association for computational Linguistics, 1994a, p. 95-100.

[13] R. Frederking, S. Nirenburg et al., "Integrating Translations from Multiple Sources within the PANGLOSS Mark III Machine Translation System", *in Proc. of the First Conference of the Association for Machine Translation in the Americas*, 1994b, p. 73-80.

[14] Sergei Nirenburg, Robert Frederking, "Toward Multi-Engine Machine Translation", Association for Computational Linguistics Morristown, NJ, USA, 1994.

[15] Tadashi Nomoto, "Multi-Engine Machine Translation with Voted Language Model", Association for Computational Linguistics Morristown, NJ, USA, 2004.

[16] Wenpeng Lu, Ruojuan Xue, "Comparative Study on Multi-systems Combination in Machine Translation", *in proc. of IEEE 3rd International Conference on Computer Science and Information Technology (ICCSIT)*, 2010, p. 308 – 312.

[17] Xiaodong HE and Mei Yang et al, "Improved Monolingual Hypothesis Alignment for Machine Translation System Combination", *ACM Transactions on Asian Language Information Processing*, Vol. 8, No. 2, Article 6, 2009.

[18] Yasuhiro Akiba, Taro Watanabe, Eiichiro Sumita, "Using Language and translation models to select the best among outputs from Multiple MT systems". *In Proc. of the 19th international conference on Computational linguistics Volume 1*, 2002, p. 1-7.

[19] Yu Chen, et al., "Multi-Engine Machine Translation with an Open-Source Decoder for Statistical Machine Translation", *in Proc. of the Second Workshop on Statistical Machine Translation*, Association for Computational Linguistics, 2007, p. 193–196.

[20] Yu Chen et al., "Combining Multi-Engine Translations with Moses", *in Proc. of the Fourth Workshop on Statistical Machine Translation*, Athens, Greece, 2009, p. 42–46.