Experiments in Cross Language Query Focused Multi-Document Summarization

Prasad Pingali, Jagadeesh Jagarlamudi and Vasudeva Varma
Language Technologies Research Centre
IIIT, Hyderabad
India.
pvvpr@iiit.ac.in, j_jagdeesh@research.iiit.ac.in, vv@iiit.ac.in

Abstract

The twin challenges of massive information overload via the web and ubiquitous computers present us with an unavoidable task: developing techniques to handle multilingual information robustly and efficiently, with as high quality performance as possible. Previous research activities on multilingual information access systems have studied cross-language information retrieval (CLIR), information extraction and factoid based question answering tasks in detail. It is believed by the research community and also previously acknowledged in a US-NSF report[Hovy *et al.*, 1999] that a cross-language query focused summarization could play a vital role in multilingual information access, as a bridge between CLIR and machine translation. Surprisingly enough, no detailed studies exist yet on the effects of cross-linguality to query focused summarization. In this paper we study the effects of adding a cross-language dimension to query focused multi-document summarization for the Telugu-English language pair. We use a cross-lingual relevance based language modeling approach to generate extraction based summary. We evaluate the system using DUC[1] 2005 dataset using ROUGE[2] metrics and compare with the mono-lingual baseline which uses relevance based language modeling in mono-lingual setting.

1 Introduction

Cross-lingual Information Retrieval (CLIR) involves the study of systems that accept queries for information in a given language L_1 and return textual documents of other language L_2, translated back into the user's language L_1 [Grefenstette, 1998]. The rapid growth and online availability of information in many languages has made this a highly relevant field of research within the broad umbrella of language processing research. Task focused workshops such as TREC[3], CLEF[4] and NTCIR[5] have focused in this area of research in the past. However, these task focused workshops and many other research projects have ignored the issues pertaining to machine translation (MT) of the retrieved results, and focus on the extensions required of traditional information retrieval (IR) to handle more than one language. Translating the search results provided by CLIR directly may be a challenging task due to the various genre of documents returned from search. Also most of the machine translation systems require syntactically well formed sentences as input in order to be able to translate documents well. Apart from these issues, machine translation systems are resource intensive programs and have problems in dealing with word sense ambiguities in a broad domain text. In this paper we discuss a cross-language query focused multi-document summarization task as a bridge between CLIR and MT. An ideal version of such a bridge module could potentially handle many of the constraints of machine translation systems mentioned above and provide text output in a way the MT system can easily handle. Such systems were previously hypothesized by natural language research community [Hovy *et al.*, 1999] but no experiments exist till date on such systems.

2 Motivation

This research activity is driven by a larger goal of enabling online information access to substantial number of Indian language speakers (about 15% of human race) who are otherwise unable to do so due to lack of understanding of English among other reasons. Research communities working on Indian languages have already been working on machine translation [Bharati *et al.*, 2002] and CLIR projects [Pingali *et al.*, 2006] which makes a good case for a query focused multi-document summarization system which can generate a readable, coherent, brief and informative query focused summary of the search results of a CLIR system.

A typical use case of such a system can be viewed as shown in Figure 1. An end user can pose a query to a CLIR system in an Indian language Language 'A'. The CLIR system can

[1] DUC - Document Understanding Conference, http://duc.nist.gov

[2] ROUGE - (Recall Oriented Understudy Gisting Evaluation) is an automatic summarization evaluation framework.

[3] TREC - Text Retrieval Conference, http://trec.nist.gov

[4] CLEF - Cross Language Evaluation Forum, http://www.clef-campaign.org

[5] NTCIR - http://research.nii.ac.jp/ntcir/

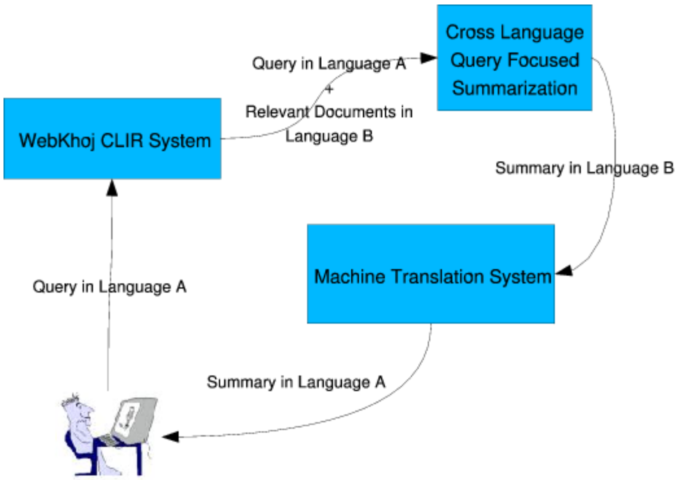

Figure 1: A Use Case for Cross Language Query Focused Multi-Document Summarization

search a huge repository of English documents and rank the relevant documents. A small set of top ranked documents along with the user's query can be input to a cross-language query focused summarizer which generates a brief and readable summary in English. The English summary can then be processed by an English to Indian language machine translation system to translate the summary back in Indian Language 'A' and provided to the end-user.

An ideal cross language query focused summarization system would drastically reduce the amount of information that needs to be translated back into the end-user's language. It would provide a syntactically well formed set of sentences in the summary to enable easy machine translation. Among other benefits such a system can also tailor its output in an easily translatable content (for example by minimizing ambiguities etc.). However this aspect of summarization was never researched before, but it would be interesting to see what could be the issues in building such a feature into summarization systems.

3 Our Contributions

Successful performance in query focused summarization tasks will usually require a system to contain a combination of IR and NLP capabilities, including passage retrieval, compression, and generation of fluent text. However, approaches like Language Modeling, Concept linkages, and Bayesian framework [Daume and Marcu, 2005; Blair-Goldensohn, 2005; Ye *et al.*, 2005; Li *et al.*, 2005; Schlesinger and Baker, 2001] provided a way to achieve good performance without involving deeper processing of text. Motivated by observation[Amigo *et al.*,] that in information synthesis tasks like this, metrics based on key concepts overlap give better results, [Jagadeesh *et al.*, 2005] discussed how representation of words in a higher dimensional semantic spaces along with relevance based language modeling can be effectively used to generate summaries. In this paper we extend the computation of metrics based on key concepts overlap across languages using cross-lingual relevance based language modeling framework adapted from similar techniques defined in a CLIR setting as in [Lavrenko *et al.*, 2002].

The main contributions of this paper can be seen as

- Extension of a relevance based language modeling technique to achieve summary extracts in a cross-lingual setting by scoring sentences.
- Definition and design of experimental setup for a cross-lingual query focused multi-document summarization task using existing monolingual summarization tasks of DUC[1] 2005 dataset.
- Testing the framework using a bilingual dictionary and HAL (Hyperspace Analogue to Language) sentence scoring technique using ROUGE evaluation for Telugu-English language pair where queries are in Telugu and the generated summary is in English.

4 Problem Statement

The problem statement is defined as to synthesize from a small set of 25-50 documents in a language L_2 that are related to a given topic, a brief, well-organized, fluent answer to a need for information given in a language L_1, that cannot be met by just stating a name, date, quantity, etc.

This task will model real-world complex cross language question answering task. Therefore the problem is different when compared with factoid based question answering tasks addressed in TREC and other similar workshops, in two aspects. The first aspect is that the answer to be generated is a descriptive one as opposed to a factoid answer and the second aspect is the corpus characteristics from which such an answer needs to be generated.

In a TREC scenario, the corpus from which answers are to be generated are large and multi-topic corpora, whereas in the problem we are dealing with here, we assume a small set of documents that are related to the user's query. This distinction would imply that the problem of Information Retrieval of the set of related documents to the query is outside the scope of the task we are addressing in this paper and the problem we are trying to address here is that of generating a summary from already retrieved relevant documents.

5 Cross-lingual Extraction Based Summarization

Extraction based summarization techniques involve ranking of textual units, usually sentences, using some scoring mechanism and picking the top scoring units and concatenating them in a certain order to generate the summary. In our approach, a cross-lingual relevance based language modeling framework is used to score sentences and generate an extraction based summary. The overall summary generation task is achieved in two steps. In the first step, called sentence scoring, the system computes query relevance score for each sentence as described in Section 5.4. The second step, called re-ranking step, the system takes the scored sentences and selects a subset of sentences which form the summary satisfying the required constraints. The re-ranker checks for the redundancy of information, computed as cosine similarity measure above a pre-defined threshold value, across the summary sentences. In future it can be extended to handle more sophisticated redundancy checking measures like (Maximum Marginal Relevance)MMR[Carbonell and Goldstein, 1998] and paraphrasing based redundancy elimination in sentences.

5.1 Cross-lingual Relevance Based Language Model

In Statistical language models[Song and Croft, 1999], documents are ranked based on the probability of producing the query from the corresponding language model of the documents. One difficulty in applying statistical language modeling to information retrieval is the sparseness of data to compute the document model. The relevance based language modeling[Lavrenko and Croft, 2001] is a significant improvement in estimating the relevance model of a document in Information Retrieval when no training data is available in the form of relevance judgments. It does not assume the query as a sample from any specific document model, instead it assumes both the query and the document as samples from an unknown relevance model R, hence it is able to overcome the problem of sparseness in the training data. In an IR task, the relevance based language model approximates $P(w|R)$, the probability of observing a word w in the documents relevant to a particular information need R, using the probability of observing the word in the context of the query $P(w|Q)$, where $Q = q_1, q_2 \ldots q_k$ is the set of query words obtained after removing the stop words from user's query. By definition, the conditional probability can be expressed in terms of the joint probability of observing some word w with the query words $q_1, q_2 \ldots q_k$. [6]

$$\begin{aligned} P(w|R) \approx P(w|Q) &= P(w|q_1, q_2 \ldots q_k) \\ &= \frac{P(w, q_1 \ldots q_k)}{P(q_1 \ldots q_k)} \end{aligned} \quad (1)$$

Conditional Sampling[Bruza and Song, 2003; Lavrenko and Croft, 2001] can be used to calculate the required joint probability $P(w, q_1 \ldots q_k)$. It assumes the query words $q1, \ldots, q_k$ to be independent of each other while keeping their dependencies on w intact.

$$P(w, q_1, \ldots q_k) = P(w) \prod_{i=1}^{k} P(q_i|w) \quad (2)$$

In a cross language setting, we extend the required term dependencies, $P(q_i|w)$, where q_i is a query word in language L_1 and w is some word being studied in language L_2, into the above expression as a joint probability of translation and post-translation query expansion obtained by the HAL model which are explained in the next sections. Therefore

$$P(q_i|w) = \sum_{j=1}^{n} P(q_i|e_j).P(e_j|w) \quad (3)$$

where each e_j is a possible translation of q_i and n is the number of possible translations for the given query term q_i. It can be observed from equation 3 that if q_i is equal and only equal to e_j (a monolingual system where L_1 is equal to L_2), we find the monolingual task to be a subset of cross-lingual task, since the term $P(q_i|e_j)$ and n become 1.

Now the translation probability $P(q_i|e_j)$ can be achieved using a bilingual lexicon and the conditional probability $P(e_j|w)$ can be incorporated into the above expression using the probabilistic interpretation of HAL model. The computation of these terms are explained in the next sections.

5.2 Calculating Translation Probability $P(e_j|w)$

Translation probability calculation has been explored in many statistical machine translation applications [Brown *et al.*, 1990; Brown and Frederking, 1995] as well as CLIR applications [Lavrenko *et al.*, 2002] in the past. Most of the applications create a statistical lexicon with translation probabilities using parallel corpora. However, in cases where parallel corpora is not available a uniform probability distribution for all possible translations can be assumed as suggested in [Lavrenko *et al.*, 2002]. In our system we use a Telugu-English bilingual lexicon assuming uniform translation probabilities for all possible translations. Though such an assumption would degrade the performance in comparison to parallel corpora based approach, it has been shown that such degradation is not much in a CLIR task [Lavrenko *et al.*, 2002]. We make use of a digitized version of a human-readable Telugu-English dictionary[7] and convert it into machine processable form. In order to handle proper names and transliteration cases of query terms we follow a set of heuristics to obtain the translation probabilities. In case any term was not found in the dictionary, the query term is transliterated into Roman text using a set of phonetic mappings across the two languages. Once the transliterations are obtained, a double metaphone algorithm [Philips, 2000] is used to obtain all the similar sounding words to the query term from the given document corpus. The words thus can be treated as all the possible translations of the given query word. This technique seemed to work very well since we are dealing with a small corpus of documents for each summary, thereby minimizing the noisy word mappings.

A set of preprocessing steps are carried out on the query terms before probabilities are calculated. In agglutinative languages such as Telugu tokenizing the query Q into query tokens $q_1, q_2 \ldots q_k$ calls for a good lemmatization module. We use an exhaustive list of affixes for Telugu in order to come up with a proper tokenization of the query. We then eliminate all the stop tokens from the generated query tokens to obtain the final list of query tokens.

5.3 Calculating $P(e_j|w)$ using Hyperspace Analogue to Language (HAL)

Motivated by the fact that the meaning of a new concept can be learnt from its usage with other concepts with in the same concept, HAL[Lund and Burgess, 1996] model constructs automatically the dependencies of a word w on other words based on their lexical co-occurrence in the context of w in a sufficiently large corpus.

[6] k is used as a temporary variable in this paper where ever necessary

[7] http://ltrc.iiit.net/onlineServices/Dictionaries/

The HAL matrix is constructed by taking a window of length k words and moving it across the corpus at one term increments. All words in the window are said to co-occur with the first word with strengths inversely proportional to the distance between them. The weights assigned to each co-occurrence of terms are accumulated over the entire corpus. That is, if $n(w, k, w')$ denote the number of times word w' occurs k distance away from w when considered a window of length K, and $W(k) = K - k + 1$ denotes the strength of this co-occurrence between the two words, then

$$\text{HAL}(w'|w) \; = \; \sum_{k=0}^{K} W(k) \, n(w, k, w')$$

The pHAL, probabilistic HAL, interpreted as, given a word w what is the probability of associating a word w' with w in a window of size K, can be expressed in terms of probability of observing w' at a distance of $k < K$ from w, as

$$\text{pHAL}(w'|w) \; = \; \sum_{k=0}^{K} P(k) \, P(w'|w, k) \tag{4}$$

where $P(w'|w, k) = \frac{n(w,k,w')}{\sum_{w''} n(w,k,w'')}$

To ensure that we obtain valid probability distribution, the constraint $\sum_{w'} pHAL(w'|w) \; = \; 1$ is imposed.

5.4 Sentence Score

From equations 1,2, 3 and 4, the relevance of a word towards the information need or the probability of observing a word w in sentences relevant to an information need can be calculated as,

$$P(w|R) \approx \frac{P(w)}{P(Q)} \prod_{q_j} P(q_j|w)$$

Assuming that the different words in a sentence are independent and removing the constant terms $c, P(Q)$, the relevance of a sentence S, can be expressed as,

$$P(S|R) \approx \prod_{w_i \in S} P(w_i) \prod_{q_j} \mathrm{P}(q_j|w_i)$$

We have retained the parameters of HAL to be same as those used in [Jagadeesh *et al.*, 2005]. The window size of 8 is used while constructing the HAL matrix. Once the sentences are scored as mentioned in the beginning of Section 5, we pick the top scoring sentences and eliminate redundant sentences using cosine similarity distance function. Finally the sentences are concatenated in the order of the publication date and position of the sentence in the documents to obtain the summary.

6 Experiments / Discussion

6.1 Experimental setup

The DUC 2005 topics in English were manually translated into Telugu by a native Telugu speaker. These topics are given as input to our system which generates a summary serving the information need represented in query. Similar to the DUC 2005 task, we restricted the length of summary to 250 words (whitespace-delimited tokens). Summaries over the size limit were truncated. No bonus was given for creating a shorter summary. While the system can generate summaries of whatever length requested by the user, in a DUC task the summary length of 250 was decided based on the following analysis. In TREC 2005 relationship task each individual question was associated with 5.6 vital nuggets on average, while on average the number of gold nuggets identified by NIST assessors were approximately 36 words long. Assuming a system is able to identify all the vital nuggets, it needs at least 200 words to answer a question. Apart this reason, the human written model summaries dataset provided by DUC is of 250 words which is another reason to choose 250 words as the length of the summary being evaluated.

An ideal summary should include (in some form or other) all the information in the documents that contributes to meeting the information need. The documents from which summary was to be generated were news articles and reports chosen from Financial Times of London and LA Times. As part of DUC 2005 a set of humans created an 250-word summaries of the cluster that meets the need expressed in each topic. These summaries are treated as model summaries against which machine generated summaries (also known as peer summaries) are evaluated using ROUGE [Lin and Hovy, 2003]. For example an input topic could be as shown in Table 1. This topic was manually translated from an English topic from DUC 2005 dataset. The answer to this information need is generated in English as shown in Table 1 which is the peer summary and gets evaluated against the model summaries.

“Who has criticized the World Bank and what criticisms have they made of World Bank policies, activities or personnel. What has the Bank done to respond to the criticisms”
ప్రపంచ బ్యాంకును ఎవరు విమర్శించారు? బ్యాంకు యొక్క విధి విధానాలపై ఏమే విమర్శలు చేశారు? వీటిపై ప్రతిస్పందించడానికి బ్యాంకు ఏమి చేసింది?
The prospect of lending to the private sector frightens some bank officials. THE World Bank is to make substantial modifications in plans for one of the world’s biggest dam projects, after accepting criticism of the project in an independent report as justified. In 1989, its first year of operation, the programme attracted the ire of the World Bank. ’ It is time for northern governments ’to face up to the fact that the IMF has failed Africa . . .’. One of the most frequent accusations is that the bank’s projects are also destroying the environment. It is as important for the Bank to admit - and learn from - such mistakes as it is to stand up for itself. ’There is also a consciousness that the Bank has to be managed. Improvement in efficiency and reduction of costs have to become an integral part of the work of the Bank. ’ The task is made urgent by the criticism of the institution and the rapidity with which the world continues to change. The World Bank has been expanding rapidly for years. The contribution of the World Bank to international lending is now marginal. It has not been a particularly happy 50th birthday for the International Monetary Fund and the World Bank. The pressure on the World Bank is rather to adopt quotas for increased lending to social sectors. The core functions of the Fund and World Bank are not, of course, their only ones.

Table 1: Box1 : Original topic in DUC 2005. Box2: Manually translated Telugu Question for the original topic Box3: English summary generated for the set of Telugu questions.

6.2 Evaluation

We evaluated the performance of the system using DUC 2005 data set. Each topic has either 4 (for 30 topics) or 9 (for the rest of 20 topics) model summaries that were written by humans, for the evaluation of peer summaries. A set of Telugu topics translated by a native speaker as mentioned above was used for generating summaries.

In DUC 2005, the evaluation of peer summaries was done both manually, for the Responsiveness, and by automatic evaluation techniques like ROUGE[Lin and Hovy, 2003]. Responsiveness was primarily measured in terms of the amount of information in the summary that actually helps to satisfy the information need expressed in the DUC topic. ROUGE evaluation framework defines multiple types of ROUGE scores designed in such a way as to suit evaluation of summaries of different lengths and characteristics. The kind of summaries being generated by our system are best evaluated using ROUGE-2 and ROUGE-SU4 metrics. This fact was reiterated in [Dang, 2005] that the automatic scores calculated using ROUGE-2 and ROUGE-SU4 correlated very well with manual evaluations with a Spearman's rank correlation coefficient of 0.95, 0.94 respectively.

Table 2 and the Figure 2 show the performance of our system "Tel-Eng-Sum" when compared to the top 10 systems that participated in DUC 2005 in terms of ROUGE-2 and ROUGE-SU4 f-measure scores. All the ROUGE scores mentioned in this paper were computed as per the ROUGE command-line parameters provided by DUC. SystemID 8 is our official system that was used in our DUC2005 participation, which uses a combination of latent semantic indexing along with HAL. It was observed that such a combination deteriorated our system performance. We have also compared the performance of HAL feature (mono-lingual baseline system), so that this would give an idea of the amount of degradation in the performance of the system due to the cross-lingual aspect. It can also be observed from Figure 3 that adding a cross-language dimension to query focused summarization task consistently degrades the ROUGE-SU4 f-measure across all the topics.

Sys ID	ROUGE-SU4	ROUGE-2
HAL feature	0.13307	
17	0.1297	0.0717
8	0.1279	0.0696
4	0.1277	0.0685
10	0.1252	0.0698
5	0.1232	0.0674
11	0.1225	0.0642
19	0.1218	0.0632
Tel-Eng-Sum	**0.12058**	**0.06048**
16	0.1189	0.0632
7	0.1189	0.0627

Table 2: Comparison of cross-lingual summarization system Tel-Eng-Sum with official scores (F-Measure) of DUC '05 participants and mono-lingual HAL feature baseline

6.3 Discussion

The problem of cross-language query focused multi-document summarization task involves many challenges covering a wide range of NLP problem domains. Apart from the

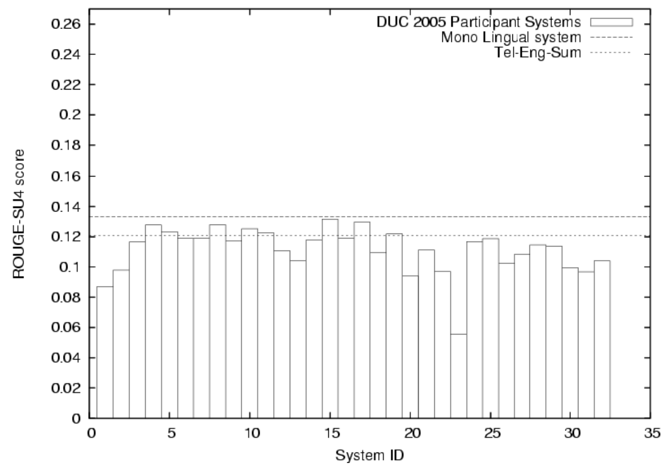

Figure 2: Telugu / English Summarization ROUGE-SU4 f-measure comparison with all the 32 systems that participated in DUC 2005

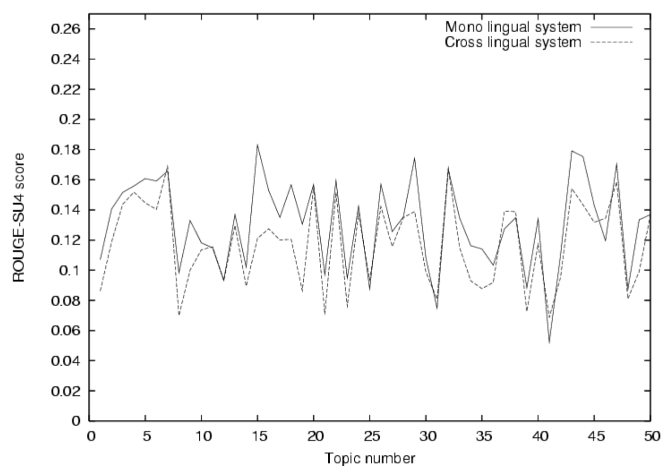

Figure 3: Telugu / English summarization ROUGE-SU4 f-measure in all the 50 topics

solutions to the problem, we realized that evaluating such systems is by itself a challenge enough that needs attention.Most of the evaluation strategies for information access systems are based on systems evaluation which can be conducted in laboratories in a controlled environement and which can eliminate user and context based biases. Inline with the same tradition, we have also used the existing monolingual system based evaluation framework being used for DUC to address the problem of cross-lingual query focused summarization evaluation. Our experimental studies show that such an evaluation framework, though not equivalent to user-based evaluations, still can give some indication of the performance of the system. However, the correlation of such an evaluation framework with actual manual evaluations has not be tested in a cross-lingual setting and is assumed based on the studies conducted in a monolingual environment. Given that the cross-language queries are manually translated from English to the language of interest and hence are representing semantically equivalent information need, such an assumption is a reasonable assumption to make.

We observed that the percentage of deterioration of performance of cross-language summarization system to it's monolingual equivalent is significantly lower when simple techniques such as a dictionary based approach are used. This finding is in contrast with other CLIR systems that use similar techniques and have significant drop in performances. We attribute the reason for such a phenomenon to the type of input data both these systems work on. In the experimental setup discussed in this paper the input set of documents are assumed to be related to the topic, which is implicitly providing document set with lower entropy and hence some biases. Despite such assumptions in the experimental setup, we feel that the results are encouraging and are suggesting that if we have had a very good CLIR system which retrieves relevant documents, the deterioration in cross-language query focused summarization is much lower. However, it would also be interesting to study the effect of deterioration in the quality of the input documents. This problem has not been studied in our experiments, as well as in DUC, since it was thought to be an IR problem rather than a summarization problem. Also we would like to point that if a better approach is used which can try to have natural language understanding capabilities, the performance deterioration may be even lower, since the system that we describe here does not make any attempt to understand the input query. The input query keywords in our system are treated to be independent of each other and the system performs purely on statistical basis. Our system is robust enough to handle unreliable resources, since the performance of the system can never go lower than the inherent information already present in the relevant set of documents which is coming from the co-ocurrence statistics of various keywords. This fact can be explained in a slightly different perspective of that of a comparison between query-focused and query-independent summarization of the set of relevant documents. We think that there would be a bottom threshold beyond which there may not be much performance drop since the input set of documents already contain some information about the topic which can counter the lack of perfect translation from the query.

The language generation aspects of the summarization system can be focused in future studies. In this system we did not focus much on this aspect, and have just eliminated approximately redundant sentences (those which have high cosine similarity with other sentences) and glue the sentences in the order they occur in the input documents.

7 Conclusion and Future Work

In this paper, we presented experiments in cross language query based multi-document summarization task for Telugu - English language pair using a relevance based language modeling framework. We extended the existing monolingual summarization technique using relevance based language modeling to a cross-language one. We calculated the translation and post-translation query expansion probabilities using a bilingual dictionary and HAL feature. We observe from the ROUGE scores that the cross-lingual system performs 85-90% compared to the monolingual baseline in ROUGE-2 and ROUGE-SU4. Relevance based modeling technique in CLIR [Lavrenko *et al.*, 2002] task for English - Chinese was previously reported to perform at about 93-97% of monolingual baseline. We plan to extend this work by studying some of the ideas mentioned in this paper, such as, customizing summary outputs for machine translation systems. We also plan to study the effectiveness of translating summaries into the source language, which is Telugu in this case. Studying the effect of query-focused summarization when compared to query-independent summarization from a set of relevant documents is also of interest of future study.

References

[Amigo *et al.*,] Enrique Amigo, Julio Gonzalo, Victor Peinado, Anselmo Penas, and Felisa Vedejo. An Emperical Study of Information Synthesis Tasks. In *Proceedings of Association for Computational Linguistics (ACL), july 2004, Barcelona*.

[Bharati *et al.*, 2002] Akshar Bharati, Rajeev Sangal, Dipti M Sharma, and Amba P Kulkarni. Machine translation activities in India: A survey. In *In the Proceedings of workshop on survey on Research and Development of Machine Translation in Asian Countries*, 2002.

[Blair-Goldensohn, 2005] S. Blair-Goldensohn. Columbia University at DUC 2005. In *Prceedings of Document Understanding Conferences*, 2005.

[Brown and Frederking, 1995] R. Brown and R. Frederking. Applying Statistical English Language Modeling to Symbolic Machine Translation. In *Proceedings of the Conference on Theoretical and Methodological Issues in MT*, pages 221–239, 1995.

[Brown *et al.*, 1990] P.F. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, J. Lafferty, R. Mercer, and P. Roossin. A Statistical Approach to Machine Translation. In *Computational Linguistics*, pages 79–85, 1990.

[Bruza and Song, 2003] Peter Bruza and Dawei Song. A comparison of various approaches for using probabilistic dependencies in language modeling. In *Proceedings of International ACM SIGIR conference on Research and development in informaion retrieval*, pages 419–420, 2003.

[Carbonell and Goldstein, 1998] Jaime G. Carbonell and Jade Goldstein. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In Alistair Moffat and Justin Zobel, editors, *International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR98*, pages 335–336, Melbourne, Australia, 1998.

[Dang, 2005] Hoa Trang Dang. Overview of duc 2005 (draft). In *Proceedings of Document Understanding Conferences*, 2005.

[Daume and Marcu, 2005] Hal Daume and Daniel Marcu. Bayesian Multi-document Summarization at MSE. In *ACL 2005, Workshop on Multilingual Summarization Evaluation*, 2005.

[Grefenstette, 1998] Gregory Grefenstette. *Cross-Language Information Retrieval*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[Hovy *et al.*, 1999] Eduard Hovy, Nancy Ide, Robert Frederking, Joseph Mariani, and Antonio Zampolli. Multilingual Information Management: Current Levels and Future Abilities. In *A report comissioned by the US NSF*, USA, 1999.

[Jagadeesh *et al.*, 2005] J Jagadeesh, Prasad Pingali, and Vasudeva Varma. A relevance-based language modeling approach to duc 2005. In *Proceedings of Document Understanding Conferences (along with HLT-EMNLP 2005), Vancouver, Canada*, 2005.

[Lavrenko and Croft, 2001] Victor Lavrenko and W. Bruce Croft. Relevance-Based Language Models. In *International ACM SIGIR conference on Research and development in Information Retrieval*, pages 120–127, 2001.

[Lavrenko *et al.*, 2002] Victor Lavrenko, Martin Choquette, and W. Bruce Croft. Cross-lingual Relevance Models. In *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 175–182, New York, NY, USA, 2002. ACM Press.

[Li *et al.*, 2005] W. Li, W. Li, B. Li, Q. Chen, and M. Wu. The Hong Kong Polytechnic University at DUC 2005. In *Proceedings of Document Understanding Conferences*, 2005.

[Lin and Hovy, 2003] Chin-Yew Lin and E.H. Hovy. Automatic Evaluation of Summaries Using N-gram Cooccurrence Statistics. In *Proceedings of 2003 Language Technology Conference (HLT-NAACL 2003), Edmonton, Canada*, 2003.

[Lund and Burgess, 1996] K Lund and C Burgess. Producing high-dimensional semantic spaces from lexical cooccurrence. In *Behavior Research Methods, Instrumentation, and Computers*, pages 203–208, 1996.

[Philips, 2000] L. Philips. The Double-Metaphone Search Algorithm. *C/C++ User's Journal*, 18(6), 2000.

[Pingali *et al.*, 2006] Prasad Pingali, Jagadeesh Jagarlamudi, and Vasudeva Varma. WebKhoj: Indian Language IR from Multiple Character Encodings. In *In the Proceedings of WWW'06*, 2006.

[Schlesinger and Baker, 2001] Judith D. Schlesinger and Deborah J. Baker. Using Document Features and Statistical Modeling to Improve Query-based Summarization. In *Proceedings of Workshop on Document Understanding Conferences, DUC01*, New Orleans, LA, 2001.

[Song and Croft, 1999] Fei Song and W Bruce Croft. A general language model for information retrieval. In *Proceedings of the eighth international conference on Information and knowledge management*, pages 316–321, 1999.

[Ye *et al.*, 2005] S. Ye, L. Qiu, and T.-S. Chua. NUS at DUC 2005: Understanding Documents via Concept Links. In *Proceedings of Document Understanding Conferences*, 2005.