

ENGLISH TO PUNJABI TRANSLATION USING HYBRID APPROACH

RICHU MITTAL AND NARESH KUMAR GARG

GZS-PTU Campus/Computer Science & Engineering, Bathinda

richumittal28@gmail.com, naresh2834@rediffmail.com

ABSTRACT: Machine Translation is a task of automatic translation in which a text from one natural language is converted to another natural language .Due to globalization language translation becomes the necessary part of the human life. So This work is an attempt to develop a Machine Translation system from English to Punjabi language which can translate English sentences in to its equivalent Punjabi language using Hybrid Approach. Hybrid Approach is a combination of various approaches direct mapping, Rule Based Approach and Example Based Approach. There is no direct Machine translation system which can convert English to Punjabi language, therefore there is a need to develop such System. In this paper a Hybrid Approach is discussed which can translate a English sentence into equivalent Punjabi language sentence with an accuracy of 90%.

KEYWORDS: Natural language processing, Direct Approach, Hybrid Approach, Example Based Approach, Rule Based Approach ,Machine Translation.

INTRODUCTION

Natural Language Processing is basically a ability of a computer system to understand human sayings. Machine Translation is a part of NLP. Machine Translation involves conversion of one natural language into another natural language. Machine translation life span is too long and still it is going on. Machine Translation attracted many researchers last two decades ago. Today's world is totally dependent on machines. Translation done with the use of a machines is a dream of mankind. Researchers use different approaches to realize this dream. The first major project in this field "Georgetown Automatic Translation" was given by IBM experts. It was a translation of Russian language sentences into English. This translation makes a revolution in the field of Machine Translation. In India research on machine translation started in eighties. The first developed system is ANGLABHARTI[8] which can perform translation from English to Indian languages then ANUBHARTI is developed (translation system of Hindi to English). According to a census there are 122 languages used by Indians and 22 as a Official languages and the top most (mostly spoken) languages are Hindi, Marathi, Bengali, Tamil, Urdu, Telgu and Punjabi . These languages are spoken by around 850 million peoples and only 10% use English. So there is a huge need of a translation system which can easily and accurately perform these kind of translations from one natural language to another natural language. Punjabi language is used as a medium of communication all over in Punjab and all work of the government of Punjab done in Punjabi but data available on web is in English so there must be a system which can perform this kind of translation, Which can convert from English to Punjabi. The proposed system is a system which can easily convert English sentences into Punjabi so that everyone can easily understand the data available on web.

LITERATURE REVIEW

K.K.Batra,G.S.Lehal, Automatic Translation System from Punjabi to English[3]for Simple Sentences in

Legal Domain: The system has been developed to translate simple sentences in legal domain from Punjabi to English. Since the structure of both the languages is different, so direct approach of translation in which word by word is translated, is not possible. So, another approach indirect approach i.e. rule based approach of translation is used. R.M.K. Sinha and A. Jain, AnglaHindi English to Hindi version of the ANGLABHARTI[5]. The System use example based as well as rule based approach and also statistics to achieve higher accuracy. Bharati A, Sangal R, Sharma D 2007 SSF[7] A machine translation system named SHAKTI was designed to produce translation from English language to Indian languages by using rule based approach and statistical approach Goyal and Lehal, Hindi to Punjabi MT system (2009)[6] Goyal and Lehal of Punjabi University, Patiala, developed a Hindi to Punjabi Machine translation system based on direct approach word-to-word translation approach with a 95% accuracy. Josan G S, Lehal G S 2008[1] A Punjabi to Hindi Machine Translation developed with the use of a direct machine translation approach and use of a bilingual dictionary that contains Punjabi words and corresponding Hindi words. Gurpreet Singh Josan and Gurpreet Singh Lehal

direct machine translation system for hindi to Punjabi (2012)[9], Developed a direct machine translation system for hindi to Punjabi based upon syntactic similarities between more or less related languages. Gurpreet Singh Josan & Jagroop Kaur, "Punjabi to Hindi statistical machine transliteration" 2011[10], developed a system for Punjabi to hindi using statistical machine transliteration approach.

RESEARCH METHODOLOGY

Proposed system Algorithm: Proposed system follows different steps

Step1: Input is entered into a system in English language (source language) sentence.

Step 2: Firstly direct mapping approach is used for the translation.

Step 3: If correct translation is not there, then translation will be done with Example based translation approach.

Step 4: If translation is carried out successfully then it will gave output in equivalent Punjabi language otherwise Rule based approach is used to perform translation.

Step 5. Exact translation is done in the Punjabi language(target language).

In Proposed system hybrid approach is used, which is a combination of following approaches :

1. Direct Mapping
2. Example Based Approach
3. Rule Based Approach

Direct Mapping

Direct Mapping is used when there is no exact translation of source language sentence to target language sentence. Then system try to find the exact match of the input string in the database. This is required in certain cases where exact output is not there after translation. eg. in case of idioms and named entities.

Eg. All's well that ends well.

The Punjabi translation of this English sentence is

ਅੰਤ ਭਲੇ ਦਾ ਭਲਾ.

After translating using some other approach sometimes exact meaning of sentence is not get. So in this case the meaning of such sentences is directly stored in the database and are retrieved later whenever required. If source line failed to translate using direct approach then system tries to solve this problem using Example Based Approach.

Example based Approach

When no exact match is found into the database for the input string using direct mapping, an example based approach is used. In example based approach a set of sample examples of English sentences are stored into the database along with their Punjabi translation. Various verbs and nouns are also to be stored into different tables which are to be used in this approach. For translation, input query is transformed in to a pattern which is compared with the patterns stored in the database. If pattern found then Punjabi translation of that particular example is extracted and stored in temporary location. From this example and input query nouns and verbs are extracted from the tables and swapping takes place between these two sentences. This can be explained as follows

For example

ਰਾਮ ਰੋ ਰਿਹਾ ਹੈ

And suppose system have stored an example in the database which is

ਸ਼ਾਮ ਖੇਡ ਰਿਹਾ ਹੈ

Its equivalent translation is

Sham is playing

Now pattern of this example and the input sentence is **Noun+is+verb+ing**

Hence pattern of source sentence match with this example pattern. To translate the source sentence just replace the noun and verb of the example translation with noun and verb of source sentence respectively. After replacing noun and verb, system generate the following output

Ram is crying.

Which is exact translation of the source sentence.

Example based approach will only work if there is corresponding pattern of source sentence is available in the example set and nouns and verbs of source sentences are available in their respectable tables. If these are not available in the tables then translation will not take place.

Rule Based Approach

If exact translation is not get using direct mapping or example based approach, then one or more of the following rules are applied to the sentence in order to get the expected output.

For example: there are various rules. Two rules are mentioned here. further rules can also be developed in order to perform another translations.

Rule 1.

[Noun][Verb]

Eg. DMCH doctor booked

If noun phrase contains single word then

Punjabi translation=noun meaning+ verb meaning

If noun phrase contains more than one words and gender of first noun is 'O'(organisation), gender of first noun is 'm' and count of 2nd noun is 's'

Punjabi translation=noun meaning +dw+ noun meaning+ verb meaning

If noun phrase contains more than one words and gender of first noun is 'O'(organisation), gender of first noun is 'm' and count of 2nd noun is 'p'

If noun phrase contains more than one words and gender of first noun is 'O'(organisation), gender of first noun is 'f' and count of 2nd noun is 's'

Punjabi translation=noun meaning+ dI+ noun meaning+ verb meaning

If noun phrase contains more than one words and gender of first noun is 'O'(organisation), gender of first noun is 'f' and count of 2nd noun is 'p'

Punjabi translation=noun meaning+ dIAW+ noun meaning+ verb meaning

Rule2.

[Noun][Verb] [Noun]

Eg. Man killed wife

meaning of first noun phrase + ny+ meaning of 2nd noun phrase+meaning of verb phrase.

RESULTS & DISCUSSION

The system is evaluated 400 sentences .some outputs are as follows. Table I show the results of the inputs, entered in the system

Table 1. System results

English Sentence(source language)	Punjabi Sentences(target language)
1. Ram is walking	ਰਾਮ ਚੱਲ ਰਿਹਾ ਹੈ
2. Sham is crying	ਸ਼ਾਮ ਰੋ ਰਿਹਾ ਹੈ
3. What is ram doing	ਰਾਮ ਕੀ ਕਰ ਰਿਹਾ ਹੈ
4. Sita is Buying	ਸੀਤਾ ਖਰੀਦ ਰਹੀ ਹੈ
5. What is he doing	ਉਹ ਕੀ ਕਰ ਰਿਹਾ ਹੈ
6. What is she doing	ਉਹ ਕੀ ਕਰ ਰਹੀ ਹੈ
7. sita is buying Vegetables	ਸੀਤਾ ਸਬਜ਼ੀਆ ਖਰੀਦ ਰਹੀ ਹੈ
8. why are you late	ਤੁਸੀਂ ਦੇਰ ਕਿਉਂ ਹੋ
9. they are watching movie	ਉਹ ਫਿਲਮ ਦੇਖ ਰਹੇ ਹਨ
10. farmer are ploughing field	ਕਿਸਾਨ ਖੇਤ ਜੋਤ ਰਹੇ ਹਨ
11. ram is playing football	ਰਾਮ ਫੁਟਬਾਲ ਖੇਡ ਰਿਹਾ ਹੈ
12. john will be playing hockey	ਜੋਨ ਹਾਕੀ ਖੇਡ ਰਿਹਾ ਹੋਵੇਗਾ
13. ram will be reading with sham	ਰਾਮ ਸ਼ਾਮ ਦੇ ਨਾਲ ਪੜ੍ਹ ਰਿਹਾ ਹੋਵੇਗਾ
14. ram will be playing football	ਰਾਮ ਫੁਟਬਾਲ ਖੇਡ ਰਿਹਾ ਹੋਵੇਗਾ

Table 2. Overall accuracy table

Parameter	Numeral Values
No. of sentences	400
Correct translations	360
System Accuracy	90%

Above TABLE II show the statistics for the results obtained by our system, table shows the overall accuracy of the system with respect to different parameters.

Table 3. Comparison with existing system

	No. of sentences	Accuracy
Previous system(News Domain)	300	84%
Proposed system(general Domain)	400	90%

Above TABLE III shows the comparison of existing system with the proposed system. Existing system works in only news domain but proposed system works in general domain.

CONCLUSION & FUTURE WORK

In this paper author represent the various machine translation approaches. Source language can be converted into target language with the use of Machine translation. Hybrid approach is used in the system to convert English language sentences into Punjabi language. Hybrid system is a combination of basically three approaches direct mapping, Example based and Rule based. Overall accuracy of the system is 90% but still there is a need of improvement. The Proposed system is a prototype of an MT system. This system can be further expanded to incorporate more features. . A lot of work is left behind this system. There are linguistic problems that cannot be handled by the system. In future, System can be upgraded to solve the linguistic problems.

ACKNOWLEDGMENT

I would like to thankful of Naresh Kumar Garg Associate Professor, GZS-PTU Campus and for his valuable guidance at every step helped me a lot in improving my work and efficiency. His supportive nature has always motivated me to work hard.

REFERENCES

- Gurpreet Singh Josan, Gurpreet Singh Lehal, "Evaluation of Direct Machine Translation System For Punjabi To Hindi", International Journal of Computer Science Issues (IJCSI), 2009.
- Vishal Goyal and Gurpreet Singh Lehal, "Evaluation of Hindi to Punjabi Machine Translation System", International Journal of Computer Science Issues (IJCSI), Volume. 4, No. 1, 2009.
- K.K.Batra,G.S, "Automatic Translation System from Punjabi to English for Simple Sentence in Legal Domain".
- K. Deep, Dr. V. Goyal "Hybrid Approach for Punjabi to English Transliteration System," International Journal of Computer Applications (0975 – 8887) Volume 28 ,No.1, August 2011.
- R.M.K. Sinha, "An Engineering Perspective of Machine Translation: AnglaBharti-II and AnuBharti-II Architectures", Proceedings of International Symposium on Machine Translation, NLP and Translation Support System(iSTRANS-2004),Tata Mc Graw Hill,New Delhi.,pp.17-19. November 2004.
- Vishal Goyal, Gurpreet Singh Lehal, "Advances in Machine Translation Systems" Volume 9 ,pp.138-150. November 2009.
- A. Bharati, R.Moona, P. Reddy,B. Sankar, D.M. Sharma, R. Sangal, "Machine Translation: The Shakti Approach", Pre-Conference Tutorial at ICON-2003.
- G V Garje , G K Kharate, "Survey of Machine Translation System in India" International Journal on Natural Language Computing (IJNLC) ,Volume 2, No.4, October 2013.
- Gurpreet Singh Josan and Gurpreet Singh Lehal, "Direct Approach for Machine Translation from Punjabi to Hindi" CSI Journal of Computing | Volume 1, No.1, 2012.
- Gurpreet Singh Josan& Jagroop Kaur, "Punjabi to Hindi statistical machine transliteration" International Journal of Information Technology and Knowledge Management, Volume 4, No. 2, pp. 459-463, July-December 2011.