Morphology based Tense Aspect Disambiguation for sentences in Telugu to English Translation

Lavanya Settipalli, Sivaiah Bellamkonda, Ramachandran Vedantham

Abstract*: *Tense, aspect and modality identification of one language and translating them to another language is a complex task in machine translation. Gaining the knowledge about tenses of a language requires complete morphology analysis of that particular Language. Native speakers of the language contain inbuilt knowledge of morphology but training the machines with this knowledge needs more effort. In this paper, we are proposing Tense, Aspect Disambiguation for the Telugu language by exploring the frequent co-occurrence of verb inflections with context words. TAD approach is to build Tense dictionary for Telugu based on the hand written rules formed by morphology analysis and then automatically tagged each sentence of test data set with the tense to which it belongs. Tagged sentences then mapped to the grammar dictionary of English while translating. Our approach had performed on text written in WX notation1 by native speakers, which contains verb-included sentences.

Index Choice: Morphology Analysis, Verb Inflection, Telugu Tense Rule Dictionary (TTRD), Tense Aspect Disambiguation (TAD).

I. INTRODUCTION

Natural Language Processing (NLP) is task of making computations for the Languages. Machine Translation (MT) which translates source language sentences that are similar in the sense as the target language, plays a crucial role in NLP where it requires so many of NLP techniques like morphological, semantic, syntactic analysis and should also achieve WSD to get better performance in translation. These analysis for morphological rich language like Telugu are more complex than the developments that were done for English and giving poor accuracy.

The Telugu language is also morph-inflected rich with GNP (gender, number, and person) and with verb inflections that represent different tenses and aspects of the language which are crucial in the syntactic and semantic representation of Telugu language sentences. There is the similarity in verb infections for different tense and their progressions and this similarity causes to ambiguity in replacing the correct tense phrase to the target Language that exactly represented as in the source language. Machine translation of these tense and aspect from source to target language and performing disambiguation is more difficult because of the differences in the tense system of the languages. However, the correct replacement of verb tenses is most important because they encode the temporal order of events in a text. Unless the tense not translated correctly, it leads to misunderstandings and confusions.

In our approach, we analyzed all these ambiguities through morphology analysis and achieved disambiguation by framing hand-written rules based on the patterns that occur frequently in the Telugu sentences that can uniquely represent a tense form.

II. LITERATURE REVIEW

Tense and aspect identification was performed and researchers previously based on the analysis of the semantic structure and temporal expressions of the sentences developed methods. This work carried out by John Lee [1] and GON G ZhengXian et al. [2] using two different approaches. John Lee developed verb tense generation for English by applying the concept of anaphoric to the tenses and identified the tense and aspect dimensions with the presence of some static prepositions that comes with the tenses and participles. This approach developed a statistical model and trained data using linear CRF and outperformed majority baseline.

Whereas in [2], they developed a classifier based tense model for the tense translation of Chinese to English language. Initially, they labeled the Chinese sentences with correct tenses and trained the data with four labels as Pr-present tense; Pa-past tense; F-future tense; UNK-unknown tense and then classification performed using multiclass SVM.

G.Pratibha et al. [7] classified the Telugu sentences, which contain no verb. They classified the sentences into different classes based the semantic structures and morphology analysis of different sentences. This work was completely based on the nouns, adjectives and their formations in a sentence. But classifying the sentences which included with verbs is more difficult with so many complications like GNP variations in verb inflection.

POS tagging for the Telugu language was presented in [3] using a morphological analyzer and a fine-grained hierarchical tag-set. POS tagging had doneby observing the word internal structure by considering lexical and semantic information along with morpho-syntactic information.

Revised Manuscript Received on December 28, 2018.

Lavanya Settipalli, Computer Applications, National Institute of Techonology, Tiruchirapalli, India.

Sivaiah Bellamkonda, National Institute of Techonology, Tiruchirapalli, India.

Ramachandran Vedantham, Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, India.

Based on this information, he formed rules for are included with verbs is more difficult with so many complications like GNP variations in verb inflection.

POS tagging for the Telugu language was presented bySrinivasuBadugu [3] using a morphological analyzer anda fine-grained hierarchical tag-set. POS tagging had done by observing the word internal structure by consideringlexical and semanticinformation along with morpho-syntactic information. Based on this information, he formed rules for morphological analyzer, which can build a syntactic parser. This syntactic parser can assign correct tags and can disambiguate many cases of tag ambiguities.

III. PROPOSED METHOD

Tense Aspect Disambiguation for Telugu language is a task of identifying the correct tense of a Telugu sentencewhich is morphologically rich, means that the Telugusentences contain various verb inflection form and structures on which the tense of a sentence depends and variesvastly. In our approach, we observed the complete morphology structure of Telugu language to achieve Tense Aspect Disambiguation. We describe the ambiguity howtense of a sentence depends on their verb inflectionsthrough the following two sentences. The sentences are taken in WX notation.

sIwarojUgudikiveVlYwuMxi
(Sitarojugudikivelthundhi/Sita goes to temple daily)
gIwarepatinuMdibadikiveVlYwuMxi
(Gita repatinundibadikivelthundhi/Gita will go to school from tomorrow)

By observing the above two sentences, verb inflection in both the sentences to the root *veVlYlYu (Velthundhi*) is similar butthey are representing different tenses. First sentence representing simple present whereas second one representing future tense. So identifying the tense of sentences asper the verb inflections only will not give the requiredresult.

In this paper, we examined the pattern of verb inflection along with a co-occurrence of a word in a sentence that can uniquely represent a particular tense or aspect. Verb inflection analysis is also useful for the identification of gender, number, and person and it is explained by the sentences

1)ninnapArXivBojanaMceSAdu(Ninnapardhivbojanamchesadu/Yesterday Pardhiv ate food) (Past Tense)

2)ninnavarRaMpadetappatikepArXiviMtikivaccesAdu (NinnavarshampadetappatikiPardhivintikivachesadu/Yesterday Pardhiv had came home before it rained) (Past perfect Tense)

In the first sentence Root: *ceyu* + inflection *Adu* with no preposition presented and with time aspect *ninna* but in the second sentence Root: *vaccu* + inflection *Adu* with preposition *appatike* presented and with time aspect *ninna*. Both the sentences have same inflection and time aspect but the presence of some preposition can change the tense of the sentence. du in the verb inflection representing that the gender, number, and person of a subject as male, single and 3rd person respectively. We analyzed all these structural patterns of Telugu sentences for different tenses and aspects and according to these patterns, we formed handwritten rules from the training data of Telugu documents and then Telugu Tense Rule Dictionary (TTRD) is developed. Two test sets each with 24000 verb contain Telugu sentences are taken to assess the performance of our approach. The overall process of our TAD approach is as described in Fig. 1.

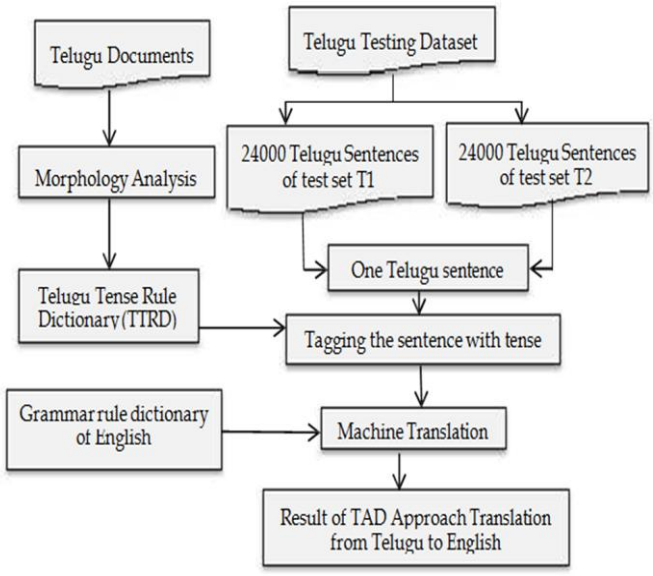


Fig.1: Overview process of TAD Approach

Telugu Language, which is a morphologically rich language, contains the words that have more than one morphology suffix. These morphological suffixes may bewith nouns or verbs. Telugu nouns are inflected for number (singular, plural), gender (masculine, feminine, andneuter) and case (nominative, accusative, genitive, dative,vocative, instrumental, and locative). The principal partsof the verb morphology are the root, the infinitive, andthe participles. There are three conjugations of Telugu verbs, each containing several classes of verbs. The fivedifferent verb forms (Present, Past, Future, and the Imperative,durative) formed with the addition of personal affixes with some particles. Generally, the main verb in the Telugu language presents at the termination of the sentence. In our exploration, we observed that the GNP (gender, number, person) problem raises the ambiguities in machine translations for many languages.

Conditions that cause ambiguity when mapping Telugu verb inflection form to English tense phrases listed below:

The Telugu language contains various verbinflection forms for different genders for a singletense in English.

- Telugu language verb inflection form itself represents the number (singular/plural) but stillthere exists some ambiguity to replace correcttense phrase of English.

  For example {*nenu*/I, *nuvvu*/you}: In Telugu, they considered as singular but in English asplural form.

- Verb form representation in the simple present for English varies according to the person of the sentence subject. Telugu verb inflection form does not give this detail.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

In our approach, to handle all these conditions, initially the sentences are grouped according to the last character, which we call it as Ex-c of the verb inflection form into six types and mapped them to GNP as in English Grammar for the gender, person and number disambiguation was presented in Table I.

Category	Ex-c	Gender	Number		Person
			Telugu	*English*	
TypeA	nu	Subjective	Singular	Plural	1stperson (I)
TypeB	mu	Subjective	Plural	Plural	1stperson (We)
TypeC	vu	Subjective	Singular	Plural	2ndPerson (you)
TypeD	du	Male	Singular	Singular	3rd person (Subject/He)
TypeE	xi	Female	Singular	Singular	3rd person (Subject/She /It)
TypeF	ru	Subjective	Plural	Plural	4thperson (they)

Table I: GNP Disambiguation In Telugu Sentences

GNP mapping itself cannot achieve disambiguation completely. Ambiguity in Machine Translation of Telugu sentence to English still exists as the inflection changes according to the gender where all those inflections represent to a single tense and a single inflection form represents different tense and aspects. These two ambiguity conditions are as presented in Table II.

Number						Tense/ Aspect	Class
Type A	Type B	Type C	Type D	Type E	Type F		
wAnu/ tAnu	wAmu/tAmu	wAvu/tAvu	wAdu/tAdu	wuMxi/tuMxi	wAru/tAru	Present Future Future perfect Future perfect continuous	Class1
unnAnu	unnAmu	unnAvu	unnAdu	uMxi	unnAru	Present continuous Past continuous Present perfect continuous Past perfect continuous	Class2
uMtAnu	uMtAmu	uMtAvu	uMtAdu	uMtuMxi	uMtAru	Future continuous	Class3
Anu	Amu	Avu	Adu	yiMxi	Aru	Past Present perfect Past perfect	Class2

Table II: Ambiguity Conditions Due To Different Verb Inflections to Classify Tense/Aspect

After the sentences had grouped as per the type, eachsentence in that type map to that particular class. However, the class of a tense still consists of ambiguity. Disambiguation of the tense class cannot solve only byanalyzing verb inflection alone. Therefore, we are considering the co-occurrence words, which can uniquely represent the tense of a sentence, and it considered as Telugu Tense Rule Dictionary (TTRD).

Telugu Tense Rule Dictionary (TTRD)

The rules are generated for the sentence to classify into tense or aspect based on the morphology analysis in the form of feature triplet as <class, co-occurrence, weight>. The feature where class and co-occurrence contain the highest weight means that they have highest likelihood had taken as the rule for that particular tense. Likelihood had calculated for the sentences from the training data and the formula to calculate the weight is as given below:

$$W(t_i, f_k) = \log\left(\frac{p(t_i|f_k)}{\sum_{j=1}^{k} p(t_j|f_k)}\right) \quad (1)$$

Where w is the weight of the feature for the tense, t_i is the tense of the sentence S_i, t_j is tense except t_i and f_k isthe k_{th} feature in the feature set. Loglikelihood estimationfor class and co-occurrences with the respective tenseshad calculated from the training data set and presented in Table III

Feature	Tense/Aspect	Likelihood
<class1, eVppudU>	Present	0.72
<class1,null>	Future	0.93
<class1, pAtiki>	Future perfect	0.97
<class1, nuMdi>	Future perfect continuous	0.82
<class2,null>	Present continuous	0.94
<class2, pAtiki>	Past Continuous	0.97
<class2, nuMdi>	Present perfect continuous	0.98
<class2, appatike>	Past perfect continuous	0.93
<class3, pAtiki>	Future continuous	0.97
<class4,null>	Past Tense	0.92
<class4, appudu>	Present perfect	0.46
<class4, appatike>	Past perfect	0.87

Table III: Likelihood Estimation For Feature And Respective Tense

Based on the maximum likelihood, the below are described as the rules for the different tenses and aspects of Telugu sentences.

- <class1, eVppudU> => Present tense
- <class4,null> => Past tense
- <class1,null> => Future tense
- <class2,null> => Present continuous
- <class2, pAtiki> => Past continuous
- <class3, pAtiki> => Future continuous
- <class4, appudu> => Present perfect
- <class4, appatiki> => Past perfect
- <class1, pAtiki> => Future perfect
- <class2, nuMdi> => Present perfect continuous

- <class2, appatiki> => Past perfect continuous
- <class1, nuMdi> => Future perfect continuous

Telugu Tense Rule Dictionary created for disambiguation of Tenses, Aspects for Telugu Language based on the generated rules, and it is as represented in Table IV.

Tense Tagging

After the dictionary of tense rules developed for Telugu language, the sentences of Telugu corpus can tagged with their particular tense. There required to preprocess the Telugu documents before going to tense tag the sentences.

class1	eVppudU null pAtiki nuMdi	Present Future Future perfect Future perfect continuous
class2	null pAtiki nuMdi appatike	Present continuous Past Continuous Present perfect continuous Past perfect continuous
class3	pAtiki	Future continuous
class4	null appudu appatike	Past Tense Present perfect Past perfect

Table IV: Telugu Tense Rule Dictionary (TTRD)

Here are the following steps that have to apply for Telugu documents before tagging process.

A. Sentence Tokenizer

Sentence tokenizing is to segment the documents into sentences, as we have to classify the sentences according to their tense. Sentence tokenizer is used outputs the sentences of the documents and then these sentences can serve for POS tagging.

B. POS Tagging

POS Tagging is the process of assigning the part of speech tags to the words. In our approach, POS tagging is required to recognize the verb part of the Telugu sentence.

C. Stemming

Stemming is the process of identifying the stem or root of a word and the inflection that added to the stem of the word. The stemming methods consider the optimal pattern of the word, which can give the correct inflection form of a stem. Our approach required stemming for verb form in a sentence to identify the verb inflection, which can be further use to analysis the tense of the sentence.

We build Algorithm1 to create the table of tagging the Telugu sentences with tense/aspect has 24000 rows and Column1 to store each sentence of test set and Column2 for tag of the respective sentence. The test set split into sentences by using sentence tokenizer for this purpose. POS tagging and stemming of a sentence to get verb and verb inflection also performed through algorithm1 to analyze the morphology structure of a sentence.

Algorithm1: TAGGING THE TELUGU SENTENCE WITHTENSE/ASPECT

Input: Telugu dataset with verb included sentences,which represent different tenses.

Output: Table of sentences and their respective tense tag.

Step 1. Split the testset intosentences using sentencetokenizer: arraySentence. Assuming that m is a number of sentences inthe dataset which is split.
Step 2. Create table tableOfTagging, which has 24000 rows and 2 columns.
Step 3. With each sentence (one sentence) in the arraySentence, do repeat i from 1 to 24000:
Step 4. S_i= arraySentence[i]
Step 5. Column1.Row[i]= S_i
Step 6. Perform POSTagging for the sentence S_i to get itsrespective verb V_i
Step 7. Perform I_i=Stemming(V_i): stemming returnsthe optimized inflection form of verb or stem
Step 8. Class = run algorithm2(I_i)
Step 9. Split this sentence into many words (or phrases) basedon ‘’ or “ ”: arrayWords. Assuming that k is a number ofwords (or phrase) of this sentence which is split.
Step 10. With each word in the arrayWords, do repeat j from 1 to k:
Step 11. if W_j is eVppudU or pAtiki or nuMdi or appatiki orappudu then W = W_j
Step 12. if Class = Class1
Step 13. if W= eVppudU then tag = Present
Step 14. else if W= pAtiki then tag = Future perfect
Step 15. else if W= nuMdi then tag = Future perfect continuous
Step 16. else tag= Future
Step 17. End of Step 12
Step 18. else if Class = Class2
Step 19. if W= appatiki then tag = Past Perfect Continuous
Step 20. else if W= pAtiki then tag = Past Continuous
Step 21. else if W= nuMdi then tag = Present perfect continuous (1)
Step 22. else tag= present continuous
Step 23. End of Step 18
Step 24. else if Class = Class3
Step 25. if W= pAtiki then tag = Future continuous
Step 26. End of Step 24
Step 27. else if Class = Class4
Step 28. if W= appatiki then tag = Past perfect
Step 29. else if W= appudu then tag = present perfect
Step 30. else tag= Past
Step 31. End of Step 27
Step 32. else tag=Invalid
Step 33. Column2.Row[i] =tag
Step 34. End of Step 10
Step 35. increment I value by 1
Step 36. End of Step 3
Step 37. Return table tableOfTagging

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Tagging of sentences with their respective tense/aspect requires identifying the class of the sentence. Therefore, we build Algorithm2 to explore and return the class of a sentence.

Class of a sentence can identify by analyzing the inflection of the verb part of the sentence for verb inflection, which sent by Algorithm1 of sentence for what we need to know class.

Algorithm2: IDENTIFYING THE CLASS OF A SENTENCE

Input: verb inflection I_i of verb V_i which belongs to the sentence S_i

Output: Class of the sentence S_i

Step 1. run Algorithm3 to get the $Type_i$ of sentence S_i that hasverb inflection I_i
Step 2. if the value of $Type_i$ is TypeA do from Step 3 to 6
Step 3. if I_i is wAnu or tAnu then $Class_i$=Class1
Step 4. elseif I_i is unnAnu then $Class_i$ =Class2
Step 5. elseif I_i is uMtAnu then $Class_i$ =Class3
Step 6. elseif I_i is Anu then $Class_i$ =Class4
Step 7. elseif the value of $Type_i$ is TypeB do from Step 8 to 11
Step 8. if I_i is wAmu or tAmu then $Class_i$=Class1
Step 9. elseif I_i is unnAmu then $Class_i$=Class2
Step 10. elseif I_i is uMtAmu then $Class_i$=Class3
Step 11. elseif I_i is Amu then $Class_i$ =Class4
Step 12. elseif the value of $Type_i$ is TypeC do from Step 13 to16
Step 13. if I_i is wAvu or tAvu then $Class_i$ =Class1
Step 14. elseif I_i is unnAvu then $Class_i$ =Class2
Step 15. elseif I_i is uMtAvu then $Class_i$ =Class3
Step 16. elseif I_i is Avu then $Class_i$ =Class4
Step 17. elseif the value of $Type_i$ is TypeD do from Step 18 to21
Step 18. if I_i is wAdu or tAdu then $Class_i$ =Class1
Step 19. Step 19: elseif I_i is unnAdu then $Class_i$ =Class2
Step 20. elseif I_i is uMtAdu then $Class_i$=Class3
Step 21. elseif I_i is Adu then $Class_i$ =Class4
Step 22. elseif the value of $Type_i$ is TypeE do from Step 23 to26
Step 23. if I_i is wuMxi or tuMxi then $Class_i$ =Class1
Step 24. elseif I_i is uMxi then $Class_i$ =Class2
Step 25. elseif I_i is uMtuMxi then $Class_i$ =Class3
Step 26. elseif I_i is yiMxi then $Class_i$ =Class4
Step 27. elseif the value of $Type_i$ is TypeF do from Step 28 to 31
Step 28. if I_i is wAru or tAru then $Class_i$ =Class1
Step 29. elseif I_i is unnAru then $Class_i$=Class2
Step 30. elseif I_i is uMtAru then $Class_i$=Class3
Step 31. elseif I_i is Aru then $Class_i$=Class4
Step 32. return $Class_i$

Identifying the class of a sentence depends on the type of the sentence. Hence we call for algorithm3 which identifies the type of the sentence by exploring the verb inflection of a sentence. The following is the algorithm code for Algorithm3.

Algorithm3: GNP DISAMBIGUATION OF TELUGU SENTENCES

Input: VERB INFLECTION I_i OF VERB V_i WHICH BELONGS TO THE SENTENCE S_i

Output: TYPE OF THE SENTENCE S_i

Step 1. Split I_i into characters: arrayChar
Step 2. Assume x as the number of characters in I_i
Step 3. Assume Ex_cas the extreme character of the inflection I_i
Step 4. Ex_c = arrayChar[x]
Step 5. if Ex_c is nu then $Type_i$ =TypeA
Step 6. elseifEx_c is mu then $Type_i$ =TypeB
Step 7. elseifEx_c is vu then $Type_i$ =TypeC
Step 8. elseifEx_c is du then $Type_i$=TypeD
Step 9. elseifEx_c is x_i then $Type_i$ =TypeE
Step 10. elseifEx_c is ru then $Type_i$ =TypeF
Step 11. return $Type_i$

IV RESULTS AND DISCUSSION

Tense Aspect Disambiguation had performed on twodifferent test sets T1 and T2 written by native speakerseach with 24000 sentences containing 2000 sentences foreach tense form. Results of tagging for the two test sets T1and T2 presented in Table V and Table VI respectively.

To implement the proposed model, we have already used Microsoft SQL Server 2008 R2 to save these test sets and save the results of tagging. Microsoft Visual Studio 2010 (C #) had used for programming to save data sets, implementing our proposed model to tag the 24,000 Telugu sentences of T1 and T2.

Tense/Aspect	#sentences in test set T1	Correctly tagged sentences	Incorrectly tagged sentences
Present	2000	1969	31
Past	2000	1984	16
Future	2000	1853	147
PresentContinuous	2000	1638	362
PastContinuous	2000	1846	154
FutureContinuous	2000	1529	471
PresentPerfect	2000	937	1063
PastPerfect	2000	1748	252
Future Perfect	2000	1152	848
Present Perfect Continuous	2000	1347	653
Past Perfect Continuous	2000	1187	813
FuturePerfectContinuous	2000	1496	504

Table V: Tagging Results On Test Set T1

Tense/Aspect	#sentences in test set T1	Correctly tagged sentences	Incorrectly tagged sentences
Present	2000	1983	17
Past	2000	1927	73
Future	2000	1896	104
PresentContinuous	2000	1712	288
PastContinuous	2000	1926	74
FutureContinuous	2000	1489	511
PresentPerfect	2000	1026	976
PastPerfect	2000	1672	328
Future Perfect	2000	1278	722
Present Perfect Continuous	2000	1284	716
Past Perfect Continuous	2000	1649	351

Tense/Aspect	#sentences in test set T1	Correctly tagged sentences	Incorrectly tagged sentences
FuturePerfectContinuous	2000	1389	611

Tabel VI: Tagging Results On Test Set T2

After tagging had performed, machine translation of the sentences had done by replacing the Telugu sentence tense with English grammar rules as per the tag of the Telugu sentence. Some sentences of our test sets, which had used to test our approach and their translation to English sentences, given by Table VII and compared our translation with the most popular translator of Google.

Translation results for the two test sets in terms of correct translation of sentences in our approach have observed and presented in Table VIII.

We have used a measure such as Accuracy (%) to calculate the accuracy of the results machine translation where accuracy is the ratio of correctly translated sentences to the total number of sentences.

Telugu Sentence	Translation Method	Translated English sentence
sIwarojugudikiveVlYlivaswuMxi	TAD Approach	Sita goes to temple daily
	Google Translator	Sita goes to church every day
rAmurojuiMtixaggarevyAyAmaMceswAdu	TAD Approach	Ramu does exercise daily at home
	Google Translator	Ramu exercises his home at home
awanuvAlYlYaMxarinikoVttAdu	TAD Approach	He beat all of them
	Google Translator	He smashed all of them
awanuakkadikiveVlYlYetappadikiakkadavarRaMpaduwuMxi	TAD Approach	When he went there, it had been raining
	Google Translator	When he goes there gets rain there
AmeVcUsepAtikiaxiveVlYlYipowuMxi	TAD Approach	It will have gone when she sees
	Google Translator	It goes away when she sees it
AmeVakkadapaniceseta ppuduvAlYlYusnehaMgAvuMdevAru	TAD Approach	They were friendly when she has worked there
	Google Translator	She was friendly when she worked there
	Google Translator	Sita goes to church every day

Table VII: Comparison of Translation between Tad And Google

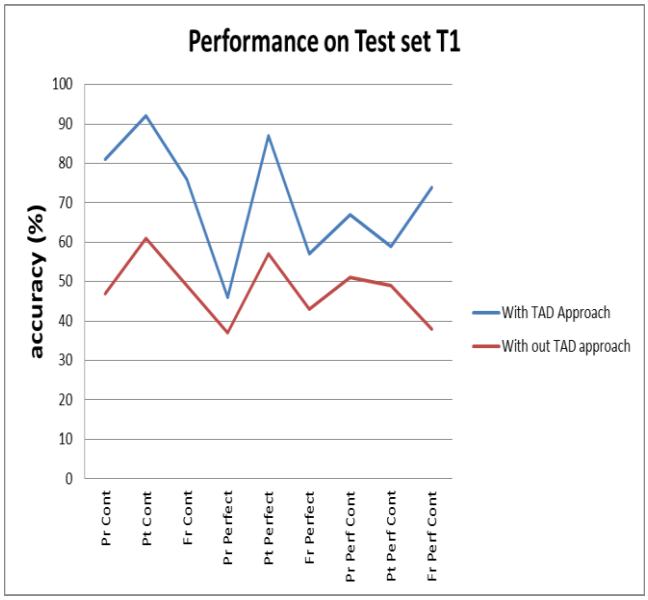
Data set	#sentences	Correctly translated
TestsetT1	24000	18686
Test set T2	24000	19112

Table VIII: Translation Results Of Tad Approach On Test Setst1 And T2

The overall accuracy of TAD approach in the translation of Telugu sentence tense to English had presented in Table IX and the results compared with Google Translator.

Data set	Method	Accuracy(%)
TestsetT1	TAD Approach	77
	Google Translator	54.74
Test set T2	TAD Approach	79.63
	Google Translator	51.27

Table IX: Overall Performance Of Tad Approach



Comparative study of our approach for every tense individually for both test sets T1 and T2 in graphical representation is as shown in Fig. 2 &3

Fig.2: Comparative study of translation accuracy with and without TAD Approach on test set T1

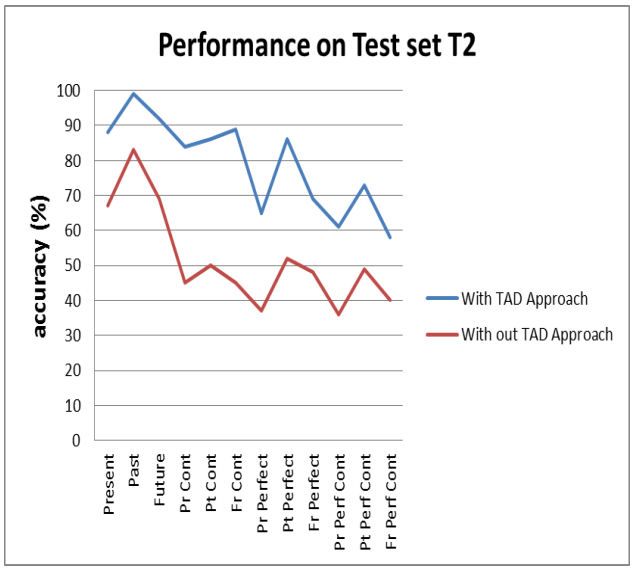


Fig.3: Comparative study of translation accuracy with and without TAD approach on test set T2

V. CONCLUSION

Tense, aspect disambiguation (TAD) model had implemented on Telugu corpus written by native speakers based on the morphology analysis of different sentences with various verb inflections. Ambiguity conditions in verb inflections had explored and the dictionary set was built to achieve disambiguation by dividing the sentences into types and classes based on verb inflection form. Disambiguation of tenses translation to the apt tense phrase according to English grammar had achieved by mapping the disambiguation conditions to English tenses by exploring the frequent co-occurrence of verb inflections with context words. Our approach wascompared with Google translator for two test sets T1 and T2 and achieved76.51%, 79.63% accuracies respectively.

Our TAD approach was limited to disambiguate tenses and aspects. Our approach can extended to modalities, negative sentences, and assertions, which are more complex than, disambiguate tenses and aspects. We also gained poor f-score for present perfect tense for which we are unable to explore the frequent co-occurrence word forpresent perfect tense of Telugu sentences in our exploration.

REFERENCES

1. John Lee, “Verb Tense Generation”, Pg No 122-130, Procedia - Social and Behavioral Sciences 27, 2011.
2. GON G ZhengX ian, ZHAN G Min, TAN ChewLim, “Classifi- er-based Tense Model for SMT”, Proceedings of COLING, 2012, pages 411–420.
3. Srinivasu Badugu, “Morphology Based POS Tagging on Telugu”, International Journal of Computer Science Issues, Pg No 181-187, Vol. 11, Issue 1, January 2014.
4. Pasquale Rullo, Veronica Lucia Policicchio, Chiara Cumbo, and Salvatore Iiritano, “Olex: Effective Rule Learning for Text Cate- gorization”, IEEE Transactions on Knowledge And Data Engi- neering, Pg No. 1118-1132, Vol. 21, NO. 8, August 2009.
5. Jisha P.Jayan, Rajeev R R, S Rajendran, “Morphological Analyser and Morphological Generator for Malayalam - Tamil Machine Translation”, International Journal of Computer Applications, ISSN: 0975 – 8887, Volume 13– No.8, PP. 15-18, January 2011.
6. G.Sindhiya Binulal, P. Anand Goud, K.P.Soman, “A SVM based approach to Telugu Parts Of Speech Tagging using SVMTool”, International Journal of Recent Trends in Engineering, Vol. 1, No. 2, PP. 183-185, May 2009.
7. G.Pratibha, Dr.Nagaratna P Hegde, “An Hybrid Approach in Classification of Telugu Sentences”, International Journal of Advanced Research in Computer Science, Volume 8, No. 5, ISSN No. 0976-5697, Pg.No. 2108-2110, June 2017.
8. W.W. Cohen and Y. Singer, “Context-Sensitive Learning Methods for Text Categorization,” ACM Trans. Information Systems, vol. 17, no. 2, pp. 141-173, 1999.
9. Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim and Sung Hyon Myaeng, "Some Effective Techniques for Naive Bayes Text Classification," IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 11, pp. 1457-1466, Nov. 2006
10. Phu Vo Ngoc, Chau Vo Thi Ngoc, Tran Vo Thi Ngoc, Dat Nguyen Duy, “A C4.5 algorithm for english emotional classification”, Springer-Verlag Berlin Heidelberg 2017.
11. W. Li, J. Han, and J. Pei, “Cmar: Accurate and Efficient Classification Based on Multiple-Class Association Rule,” Proc. First IEEE Int’l Conf. Data Mining (ICDM), 2001.
12. Jonathan J Webster and Chunyu Kit, “Tokenization as the initial phase in NLP”, In Proceedings of the 14th conference on Computational linguistics-Volume 4. Association for Computational Linguistics, PP. 1106–1110, 1992.
13. Nujaree Sukasame, Sathaporn Kantho, PenneeNarrot, “A study of errors in learning English Grammatical structures on Tenses of MatthayomSuksa 4 Students of The Demonstration School, KhonKaen University”, Procedia - Social and Behavioral Sciences 116, PP. 1934 – 1939, 2014.
14. Dinesh Kumar, Gurpreet Singh Josan, “Part of Speech Taggers for Morphologically Rich Indian Languages: A Survey”, International Journal of Computer Applications, ISSN: 0975 – 8887, Volume No.5, PP. 1-9, September 2010.
15. Aparna Varalakshmi, Atul Negi, Sai Krishna, “DataSet Generation and Feature Extraction for Telugu Hand-Written Recognition”, International Journal of Computer Science and Telecommunications, Volume 3, Issue 2, PP. 57-59, February 2012.
16. Phani Chaitanya Vempatya, Satish Chandra Prasad Nagallaa, “Automatic Sandhi Spliting Method for Telugu, an Indian Language”, Procedia - Social and Behavioral Sciences 27, PP.218–225, 2011.
17. V. Suresh, M.S. Prasad Babu, “Clause Boundary Identification for Non-Restrictive Type Complex Sentences in Telugu Language”, International Journal of Advanced Research in Com- puter Science, ISSN: 0976-5697, Volume 7, No. 4, PP. 6-10, July-26 August 2016.
18. Neepa Shah, Sunita Mahajan, “Efficient Pre-Processing for Enhanced Semantics Based Distributed Document Clustering”, International Conference on Computing for Sustainable Global development, PP. 338-343 , 2016.
19. Shahin Vaezi, Mehrasa Alizadeh, “How learners cope with English tenses: Evidence from think-aloud protocols”, Procedian- Social and Behavioral Sciences 29, PP. 986 – 993, 2011.
20. Ye, Y. and Zhang, Z., “Tense tagging for verbs in cross-lingual context: A case study”, Natural Language Processing–IJCNLP, pages 885–895, 2005.
21. F. Sebastiani, “Machine Learning in Automated Text Categorization,” ACM Computing Surveys, vol. 34, no. 1, pp. 1-7, 2002.
22. D. Isa, L. H. Lee, V. P. Kallimani and R. RajKumar, "Text Document Preprocessing with the Bayes Formula for Classification Using the Support Vector Machine," IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 9, pp. 1264-1272, Sept. 2008.
23. Himanshu Agarwal and Anirudh Mani, “Part of Speech Tagging and Chunking with Conditional Random Fields”, In Proceedings of NLPAI Ma-chine Learning Workshop on Part of Speech Tagging and Chunking for Indian languages, IIIT Hyderabad, Hyderabad, India, 2006.
24. Asif Ekbal and Samiran Mandal, "POS Tagging using HMM and Rule based Chunking", In Proceedings of International Joint Conference on Artificial Intelligence Workshop on Shallow Parsing for South Asian Languages, IIIT Hyderabad, Hyderabad, India, 2007.
25. Sarinnapakorn and M. Kubat, "Combining Subclassifiers in Text Categorization: A DST-Based Solution and a Case Study," IEEE Transactions on Knowledge and Data Engineering, vol. no. 12, pp. 1638-1651, Dec. 2007.
26. Z. Wang, G. Xu, H. Li and M. Zhang, "A Probabilistic Approach to String Transformation," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 5, pp. 1063-1075, May 2014.
27. Dalal MK, Zaveri M, “Automatic text classification: a technical review”, International Journal on Computer Applications, Volume 28, No. 2, ISSN: 0975-8887, 2011.
28. Y. Takano, "Coordination of Verbs and Two Types of Verbal Inflection," in Linguistic Inquiry, vol. 35, no. 1, pp. 168-178, Jan. 2004.
29. K. Uchimoto, K. Takaoka, C. Nobata, A. Yamada, S. Sekine and H. Isahara, "Morphological analysis of the corpus of spontaneous Japanese," IEEE Transactions on Speech and Audio Processing, vol. 12, no. 4, pp. 382-390, July 2004.
30. Jingnian Chen, Houkuan Huang, Shengfeng Tian, Youli Qu, “Feature selection for text classification with Naïve Bayes”, Expert Systems with Applications, Vol 36, PP. 5432–5435, 2006.
31. Frank, E., Bouchaert, R. R., “Naive Bayes for text classification with unbalanced classes”, In Proceedings of the 10th European conference on principles and practice of knowledge discovery in databases, pp. 503–510. Berlin: Springer, 2006.

32. A. C. R. Tsai, C. E. Wu, R. T. H. Tsai and J. Y. j. Hsu, "Building a Concept-Level Sentiment Dictionary Based on Commonsense Knowledge," IEEE Intelligent Systems, vol. 28, no. 2, pp. 22-30, March-April 2013.
33. Kim, S., Han, K., Rim, H., Myaeng, S., "Some effective techniques for Naïve Bayes text classification", IEEE Transactions onKnowledge and Data Engineering, vol18 No.11, PP.1457–1466, 2006.
34. J Steven and DeRose, "Grammatical Category Disambiguation by Statistical Optimization", Computational Linguistics, Vol. 14, No. 1, pp 31–39, 1988.
35. D. G. Lee and H. C. Rim, "Probabilistic Modeling of Korean Morphology," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 5, pp. 945-955, July 2009.
36. N. L. Bhamidipati and S. K. Pal, "Stemming via Distribution-Based Word Segregation for Classification and Retrieval," IEEETransactions on Systems, Man, and Cybernetics, vol. 37, no. 2, pp. 350-360, April 2007.