Dialogue Act Recognition for Text Based Sinhala

Sudheera Palihakkara[1], Dammina Sahabandu[2], Chamika Kasun[3], Ahsan Shamsudeen[4], Surangika Ranathunga[5]

Department of Computer Science and Engineering
University of Moratuwa
Sri Lanka
[1]sudheera.10@cse.mrt.ac.lk, [2]dammina.10@cse.mrt.ac.lk, [3]chamika.10@cse.mrt.ac.lk, [4]ahsan.10@cse.mrt.ac.lk,
[5]surangika@cse.mrt.ac.lk

Abstract—**This paper is concerned with the classical machine learning approaches to the task of Dialogue Act Recognition for text based Sinhala. A new annotated corpus for Sinhala language is built along with a tag set. We performed an evaluation based on features identified in utterances. Evaluation is performed on features that used in other studies as well as newly identified features for Sinhala. Using the best performing feature set we have performed an evaluation for effectiveness of classifiers. Considering the test results on features, we identified the best performing feature set for Sinhala language. Considering the result of classifier test, we identified the best performing classifier for Sinhala language.**

Keywords-classification; corpus; dialogue act recognition; Sinhala

I. Introduction

To understand a spontaneous dialogue, it is important to model and automatically identify the structure of that dialogue, because it will make it easier to get a better interpretation of that spontaneous dialogue. How to model a spontaneous dialogue precisely is still an open issue, though some of the specific characteristics for modeling a spontaneous dialogue have already been identified. Among these clearly identified characteristics, "Dialogue Acts" hold an important place.

The process of identifying the Dialogue Acts (DAs) for a particular language, consists of fixed set of steps [1]. That process is independent from the language we use for the Dialogue Act Recognition. The first and foremost step of the dialogue act recognition procedure is to identify the set of DA labels that is relevant for the task. After that, relevant informative features have to be computed from the speech signal. That is a very critical step since the accuracy of identifying the Dialogue Acts heavily depends on the identified feature set. Then DA models will be trained on these identified features set. Apart from DA recognition, the segmentation of the dialogue into utterances needs to be carried out independently, or alternatively realized during the recognition step with joint DA recognition and segmentation models.

This paper presents our work on dialogue act recognition for Sinhala using textual information extracted from the corpus. Next section describes about the previous studies performed on this area. An overview of the available corpora and Sinhala corpus we built is given in section III. In section IV we will talk about the selection of the tag set for the Sinhala corpus. Feature selection and the results of the test we performed are described in the section V. Various classifiers and the performance analysis of them are described in section VI. We conclude with some discussion and the conclusion of our study.

II. Related Work

There are no prior studies carried out for dialogue act recognition for Sinhala language, Alternatively a considerable amount of research carried out for other major languages including French [1], Czech [2] and Germen [3].

Most of studies have reused an existing corpus such as SWITCHBOARD [4] and VERBMOBIL [5] rather than building one. A fine set of dialogue acts are defined as the next step based on the DAMSL [6] tag set proposed by James et al. In [7], [8] Stolcke et al defined 42 dialogue act tag classes for SWITCHBOARD corpus. Usually many studies tend to reduce the defined classes to a much broader, smaller number of classes.

Dialogue act classes can be recognized using following two major models,

1. Language models
2. Prosodic models

The first type of models, uses the word related attribute of an utterances such as word sequence, specific words etc. probabilistic language models such as n-gram [7] classification trees [9] or neural networks [8], [10]. Prosodic models takes the audio signal related information into account. Energy and speaking rate is taken into consideration by defining attributes of prosodic information such as amplitude and timing statistics [11], [12] and [13].

III. Corpora

In literature we identify several standard corpora which are commonly used in the studies related to dialogue act recognition. Table 1 includes a summary of some of those widely adapted corpora.

TABLE I. Available Corpora.

Corpus	Utterance Count	Dialogue Count	Domain
SWITCHBOARD[4]	223 606	1155	Telephone conversations
VERBMOBIL[5]	3117	168	Spoken conversations
ICSI MEETING RECORDER[14]			Multi party meetings
MAPTASK[15]	26 621	128	Dialogues

As there is no such standard corpora existing for Sinhala language, we were required to build a new standard corpus to proceed with our study. Our requirement was to build a corpus (Sanwada corpus) that contains Sinhala textual conversations. So our solution was to extract conversational utterances in Sinhala movie subtitle files and manually segment them in order to build a large standard Sinhala corpus within a short period of time. When selecting subtitle files, we manually selected the movies that contain long continuous conversations and avoided the movies that has many scene changes and short conversations. At the end, Sanwada corpus comprised of 60,000 Sinhala conversational utterances extracted from 64 movies.

IV. Selecting the Dialogue Act Tag Set

The Dialogue acts are the basic building blocks for the process of spoken language understanding in human conversations. So selecting an appropriate dialogue act tag set is the crucial first step in processing conversational speech. This heavily depends on the language, culture and the context of the target application/task. Kral [1] identifies the following three requirements as the requirements for selection of tag set. These apply for any language and the any context that we are taking into account for DA recognition.

1. The DA labels should be generic enough to be useful for different tasks, or at least robust to the unpredictable variability and evolution of the target application.
2. The DA labels must be specific enough to encode detailed and exploitable characteristics of the target task.
3. The DA labels must be clear and easily separable, in order to maximize the agreement between human labelers.

Although, as the above three rules explain, the selection of a tag set heavily depends on the context of the target application/task, there are some tag sets which are usually used as common base lines for most tasks. In a study, what usually happens is, first these common tag sets are studied and then target specific DA tag sets specific for the context are derived. DAMSL tag set [6], SWBD-DAMSL tag set [16], Meeting Recorder tag set [14] and VERMOBIL tag set [3] are some of the most commonly adapted tag sets which are based on the widely used standard corpuses.

Considering the Sinhala language, although there are very few similarities between Sinhala and English they can be consider as two different languages with different sets of language characteristics. For example, in English it is an easy task to categorize the utterances into the categories, commands, orders and requests, because there is significant difference between command utterances, order utterances and request utterances. For example, in a request utterance the word "please" is commonly used while in an order it is rarely used. But with the native language characteristics of Sinhala it is hard task to find the separation between these three categories with only textual and lexical information given. So the best thing to do is combine these three categories into a single dialogue act.

To select the appropriate dialogue act tag set for the study we have followed three major steps. As the first step, we have adapted the SWBD-DAMSL tag set and have done an analysis on that tag set with respect to the other tag sets used in related studies to identify the most widely used dialogue act tags, because the number of tags that we can use for the study heavily depends on the size of the corpus and it will be almost impossible to classify utterances into the dialogue acts which occurs very rarely. So even though Switchboard corpus (223,606 utterances) used 42 major classes of dialogue acts, for our study (selected corpus size 10,000 utterances) it was suitable to use less than 20 major dialogue act classes.

So as the second step we have selected the 20 tags according to the results of the above mentioned analysis and have tried to label a sample corpus of 5000 utterances to measure the completeness and the suitability of that tag set considering our context and the characteristics of Sinhala language. At the end of the experiment we have remove five dialogue act tags due to rare occurrence rate and introduced a new tag and finalized the tag set. TABLE II displays the final tag set used to label the Sanwada corpus.

TABLE III. Selected Dialogue Act Tag Set.

English Preposition	Corresponding Sinhala Prepositions
Statement	48.51%
Yes-No Question	12.87%
Request/Command/Order	10.23%
Open Question	9.78%
Back-channel/Acknowledge	7.39%
Conventional Opening	2.58%
Backchannel Question	2.31%
No Answer	1.42%
Yes Answers	1.36%
Apology	1.33%
Thanking	0.75%
Opinion	0.44%
Abandoned/Uninterpretable/Other	0.44%
Conventional Closing	0.31%
Expressive	0.17%
Reject	0.11%

A. Dialogue Act Labeling Task

After defining the appropriate dialogue act tag set the next task was to label the corpus using those tags, we have assigned four independent annotators to tag the part of the Sanwada corpus. After that we calculated the Fleiss kappa [17] value to measure the inter-annotator agreement and we got 82% of agreement between the annotator. There were utterances that have been tagged with different tags by four annotators. So then we did retagged those utterances manually under the supervision of a Sinhala linguist.

V. Features

A. Features identified for other languages

Akker and Schulz [18] identify features as the input given for the classifier, as a vector for each word in the utterance. Features can be extracted from the word itself, timing and the prosodic information. The features in an utterance can be grouped into 5 major categories.

1. Time Related Features
2. Word Related Features
3. Prosodic Features
4. Online Features
5. Other Features

Time related and prosodic features are applicable only when the audio information of the utterances are available. Since our work is focused on the textual context we considered only the word related and online features. Next follows a brief description about word related, online and other features.

Words can act as features themselves. These are categorized under the word related features. ***Current word***, ***next word*** and ***previous word*** are 3 most commonly used features. Since most classifiers cannot deal with Strings directly, most of the time in the context, the feature of the string is converted into a nominal feature for each word. Using the same procedure, next word and previous word features can be derived from the same utterance.

Apart from above features, there are 4 other features related to segmentation of the utterances. ***Number of words in previous segment*** feature is self-explanatory. ***Distance to the last segment*** feature is the number of words from the end point of the last segment. ***Relative position of word inside the segment*** and ***Time interval of current word to last segment*** are other segmentation related features.

Rosset and Lamel [19] used a feature-vector consisting of *Speaker Identity, Number of utterances* and *First two words.* Lendvai [20] opted for not using the last DA tags as a feature for the current utterance, as it will introduce a cumulative error. *Utterance type, Presence/absence of Wh-Question* and *Subject type* were used as features in *Andernach* [21]. And also, two interesting features *1st verb type* and *2nd verb type* because of their potential of informatively on kinds of agents and actions. Similar use of above mentioned two features can be seen in other context [22] as *grammar patterns.*

B. Features identified for Sinhala

Considering above discussed categories, we have identified 10 major features that can be used for classifying Sinhala utterances. Next follows all the major features identified throughout the project.

1. **Cue Phrases :** presence of connective expressions
2. **Number of words in the segment :** self-explanatory
3. **Bigrams/Trigrams of words:** Adjacent two words in an utterance is considered as a bigram, likewise trigram is adjacent three words. We identified unigram is ineffective for Sanwada corpus during preliminary experiments. Ivanovic [23] described how unigrams worked well on live chat messages.
4. **Previous DA :** The dialogue act of the previous utterance
5. **Verb of the Sentence :** self-explanatory
6. **Punctuation marks:** The appearance of the question mark, exclamation mark, Full stop, etc. in the utterance.
7. **Grammar pattern :** The Sinhala grammar pattern(s) of the sentences in the utterance
8. **Last word of the utterance :** self-explanatory
9. **Bag of words for each tag:** For each tag the most frequent words appear in the training set of utterance.
10. **End letter of the last word of the sentence is letter ‘ද’:** The presence of special letter ‘ද’

C. Feature Selection Results

The idea of the experiment is to identify the most contributing features for classifying and the most effective combinations of the features. There were 21 features which we reduced to 9, based on the performance evaluation, in order to test with different combinations of features. Because with 21 features, it is computationally expensive to go through every possible combinations of 21 features. So we filtered out the less contributing features.

We used WEKA [24] Java library for classification. To achieve above described task we used *InfoGain Attribute Evaluator* of WEKA and obtained the *infogain* value. TABLE III displays the results.

TABLE IIIII. INDIVIDUAL FEATURE PERFORMANCE.

Rank	Feature	Infogain
1	*Punctuation marks*	*0.71*
2	*Last word of the utterance*	*0.60*
3	*Trigrams/Bigrams*	*0.31*
4	*End letter of the last word of the sentence is letter ‘ද’*	*0.30*
5	*Verb of the Sentence*	*0.24*
6	*Number of words in the segment*	*0.18*
7	*Bag of words for each tag*	*0.18*
8	*Cue Phrases*	*0.17*

The J48 classifier has been used to classify corpora with different combinations taken from above features. First we selected few features based on the *infogain* value and expanded the feature set according the F-measure value. And the resulting best performing subset of features is displayed on TABLE IV. The feature set achieved F-measure value of **0.7469** with a precision **0.7498** and recall **0.7653**.

TABLE IV. BEST PERFORMING FEATURES.

Feature
Punctuation marks
Last word of the utterance
End letter of the last word of the sentence is letter ‘ද’
Verb of the Sentence
Cue Phrases

VI. CLASSIFIERS

This section simply describes the function of each of the classification techniques available in the context. They can be broadly classified as probabilistic or statistical, decision tree, support vector machines, rule based and artificial neural network based classifiers. We have described some of available classifiers here and the results of the performance test we performed against the selected feature set above.

The Naive Bayes Classifier is a simple probabilistic classifier based on the Bayesian theorem which calculates a set of probabilities by counting the frequency and combinations of values with strong independence assumption among features. **Logistic regression** is another probabilistic statistical classification model which uses regression analysis for the classification process. **Simple logistic** model is the simplest regression model which uses linear regression. There a lot of classifiers which are based on the decision tree structure such as **J48, Simple Cart, REPTree** and **LMT**. J48 classifier is a predictive machine learning model that decides the target value of a testing data on various attribute values of the built decision tree [25]. Here, the attribute with the highest normalized information gain is chosen to make the classification decision. **REPTree** uses information gain and prunes it using reduced-error pruning. A **logistic model** tree is a standard decision tree which combines logistic regression functions at the leaves. **Radial basis function network** classifier is a type of an artificial neural network based classifier and it uses radial basis functions as the activation function of the neural network. **Decision tables** are simple supervised classifiers which group rules that have similar conditions and actions, and help to spot problems such as overlaps and gaps among the rules. **Sequential minimal optimization** is an iterative algorithm widely used to train support vector machines. **Hyperpipes** classifier contains all points for each and every category, which means it essentially records the attribute bounds observed for each category and classifies test instances according to the category that "most contains the instance".

The test was evaluated using the F-measure for each classifier on the same feature set. The TABLE V contains the results.

TABLE V. CLASSIFIER PERFORMANCE.

Classifier	Recall	Precision	F-measure
J48	*0.77*	*0.75*	***0.75***
RandomForest	*0.76*	*0.75*	***0.74***
REPTree	*0.77*	*0.77*	***0.74***
PART	*0.75*	*0.74*	***0.73***
LMT	*0.75*	*0.74*	***0.73***
DecisionTable	*0.75*	*0.76*	***0.72***
Logistic	*0.71*	*0.68*	***0.66***
SimpleLogistic	*0.71*	*0.69*	***0.66***
NaiveBayes	*0.67*	*0.61*	***0.63***
SMO	*0.69*	*0.56*	***0.61***
HoeffdingTree	*0.64*	*0.41*	***0.5***
DecisionStump	*0.64*	*0.41*	***0.5***

VII. CONCLUSION

In this paper we discussed about adapting the classification techniques to Sinhala Language. We built a corpus using Sinhala movie subtitles, and defined suitable dialogue act tag sets for Sanwada corpus based on the results of a few tests performed on the corpus. The experiments done on Sanwada corpus for recognizing dialogue acts obtained reasonable test results and showed that various word related and online features can be used for Sinhala dialogue act recognition. Our work was carried out using a relatively small number of tag sets and features, so despite the fact we achieved reasonably good results, the performance can be increased further more. The feature selection test explored new ways of extracting information from the utterances and we identified a best performing feature set for the Sinhala Language. The classifier tests revealed that most of the classifiers perform well with the Sinhala corpus without any classifier parameter tuning. We reached to 76.5317% accuracy of dialogue act tagging with J48 classifier in WEKA.

As future work, we suggest taking higher level information as prosody in to the picture and defining features on it. Classifier optimization is another aspect we have not covered in our study.

REFERENCES

[1] Král, Pavel, and Christophe Cerisara. "Dialogue act recognition approaches."*Computing and Informatics* 29.2 (2012): 227-250.

[2] Kral, Pavel, Christophe Cerisara, and Jana Kleckova. "Combination of classifiers for automatic recognition of dialog acts." *Interspeech*. 2005.

[3] Klein, Alexandra, Elisabeth Maier, Ilona Maleck, Marion Mast, and Joachim Quantz. Dialogue acts in VERBMOBIL. Univ., 1995.

[4] Godfrey, John J., Edward C. Holliman, and Jane McDaniel. "SWITCHBOARD: Telephone speech corpus for research and development." Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on. Vol. 1. IEEE, 1992.

[5] Kurematsu, Akira, Youichi Akegami, Susanne Burger, Susanne Jekat, Brigitte Lause, Victoria MacLaren, Daniela Oppermann, and Tanja Schultz. "VERBMOBIL dialogues: multifaced analysis." In INTERSPEECH, pp. 712-715. 2000.

[6] Allen, J., and G. Mark. "Core. 1997. Draft of DAMSL: Dialog Act Markup in Several Layers." (2013).

[7] Stolcke, Andreas, et al. "Dialog act modeling for conversational speech." AAAI Spring Symposium on Applying Machine Learning to Discourse Processing. 1998.

[8] Stolcke, Andreas, et al. "Dialogue act modeling for automatic tagging and recognition of conversational speech." Computational linguistics 26.3 (2000): 339-373.

[9] Mast, Marion, et al. "Automatic classification of dialog acts with semantic classification trees and polygrams." Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing. Springer Berlin Heidelberg, 1996. 217-229.

[10] Andernach, Toine, Mannes Poel, and Etto Salomons. "Finding classes of dialogue utterances with kohonen networks." ECML/MLnet Workshop on Empirical Learning of Natural Language Processing Tasks. 1997.

[11] Shriberg, Elizabeth, et al. "Can prosody aid the automatic classification of dialog acts in conversational speech?." Language and speech 41.3-4 (1998): 443-492.

[12] Mast, Marion, et al. "Dialog act classification with the help of prosody." Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on. Vol. 3. IEEE, 1996.

[13] Wright, Helen, Massimo Poesio, and Stephen Isard. "Using high level dialogue information for dialogue act recognition using prosodic features." ESCA Tutorial and Research Workshop (ETRW) on Dialogue and Prosody. 1999.

[14] Shriberg, Elizabeth, et al. The ICSI meeting recorder dialog act (MRDA) corpus. INTERNATIONAL COMPUTER SCIENCE INST BERKELEY CA, 2004.

[15] Anderson, Anne H., et al. "The HCRC map task corpus." Language and speech34.4 (1991): 351-366.

[16] Jurafsky, Dan, Elizabeth Shriberg, and Debra Biasca. "Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual." Institute of Cognitive Science Technical Report (1997): 97-102.

[17] Fleiss, J. L. "Statistical methods tor rates and proportions." Nueva York: Wiley8 (1981).

[18] op den Akker, Harm, and Christian Schulz. "Exploring features and classifiers for dialogue act segmentation." Machine Learning for Multimodal Interaction. Springer Berlin Heidelberg, 2008. 196-207.

[19] Rosset, Sophie, and Lori Lamel. "Automatic Detection of Dialog Acts Based on Multi-level Information Ý." (2004).

[20] Lendvai, Piroska, Antal van den Bosch, and Emiel Krahmer. "Machine learning for shallow interpretation of user utterances in spoken dialogue systems." Proc. of EACL-03 Workshop on Dialogue Systems: interaction, adaptation and styles of management. 2003.

[21] Andernach, Toine. "A machine learning approach to the classification of dialogue utterances." arXiv preprint cmp-lg/9607022 (1996).

[22] Jurafsky, Daniel, et al. "Lexical, prosodic, and syntactic cues for dialog acts." Proceedings of ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers. 1998.

[23] Ivanovic, Edward. Automatic instant messaging dialogue using statistical models and dialogue acts. University of Melbourne, Department of Computer Science and Software Engineering, 2008.

[24] Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter 11.1 (2009): 10-18.

[25] Patil, Tina R., and Mrs SS Sherekar. "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification." International Journal Of Computer Science And Applications 6.2 (2013).