Enhance Features in URDU.KON-TB

Saima Munir
Department of Computer Science & Information Technology, University of Sargodht,40100 Sargodha , Pakistan

Keywords: URDU.KON-TB, Converted ability, Enhance features and MaltParser.

Abstract: In this paper, we enhance features are the head & dependent relationship and functional tagset is marked by dependency grammar rules in URDU.KON-TB for increase the accuracy. The SSP and SSS tagset are using. In this way, we conduct one experiment with six different feature models using MaltParser. First we check converted ability of URDU.KON-TB in domain of dependency parsing through conversion, so that's why we need to proposed formula and defined rules.

1 INTRODUCTION

We are using a CONLL format data and make a computational model in which enhance feature of URDU.KON-TB. So our research objectives/tasks are as following:

- Check converted ability of URDU.KON-TB in domain of dependency parsing through conversion, so that's why we need to proposed formula and rules.
- Enhance features of head dependent relationship in URDU.KON-TB (Munir et al., 2017).
- The Functional tagset is marked by dependency grammar rules (Munir et al., 2017).
- The aim to check of increasing the feature in model is helpful to increase the accuracy.

2 LITERATURE REVIEW

In 2017, Munir et al. present evaluation of URDU.KON-TB in the dependency parsing domain with three types of tagset, the semi-semantic POS (SSP), semi-semantic Syntactic (SSS) and Functional (F) tagset (Abbas, 2014).
They were proposed conversion and defined 7 rules to extract data in CONLL format of MaltParser from URDU.KON-TB. The suitability, compatibility and usability of data also measured in the dependency parsing domain. To make the data compatible, few assumptions are taken. They have performed eight experiments with six different feature models and converted 80% training (data using for train MaltParser) and 20% testing data (using for test MaltParser and check performance. Test dataset has never been used in training) contains 25 sentences with average length of 15 words. An assumption based enhancement by adding Head information showing in figure 1. They get 49% accuracy with SSP and SSS tagset usable and suitable in dependency parsing domain (Munir 1 et al., 2017), (Ali et al., 2010), (J.nivre, 2006) and (Abbas, 2014).

ID	FORM	-	POSTAG	CPOSTAG	-	HEAD	DEPREL	-	-
1	حامد	-	N.PROP	KP.ERG	-	0	KP.ERG.SUB	-	-
2	نے	-	CM	KP.ERG	-	0	KP.ERG.SUB	-	-
3	شیر	-	N	KP.ACC	-	0	KP.ACC.OBJ	-	-
4	کو	-	CM	KP.ACC	-	0	KP.ACC.OBJ	-	-
5	جنگل	-	N.SPT	KP.SPT	-	0	KP.SPT.MODF	-	-
6	میں	-	CM	KP.SPT	-	0	KP.SPT.MODF	-	-
7	بندوق	-	N	KP.INST	-	0	KP.INST.MODF	-	-
8	سے	-	CM	KP.INST	-	0	KP.INST.MODF	-	-
9	مارا	-	V.PERF	VCMAIN	-	0	VCMAIN.Root	-	-

Figure 1: Enhanced with head information

Munir, S.
Enhance Features in URDU.KON-TB.
DOI: 10.5220/0009772800820085
In *Proceedings of the 1st International Conference of Computer Science and Renewable Energies (ICCSRE 2018)*, pages 82-85
ISBN: 978-989-758-431-2

3 CONVERTED ABILITY

In this section, we are going to check converted ability of URDU.KON-TB in domain of dependency parsing through conversion. The aim is claim of converted ability another domain with conversion.

So that's why, we proposed formula of converted ability is convert able tagset another domain divided by total number of tagset in Treebank and gets more 60% showing in figure 2 in which total number of tagset in URDU.KON-TB is 66 and total number of usable or able tagset in another domain is 48 according to result of research work by 2017, Munir et al. The percentage of converted ability is 72.72% using formula.

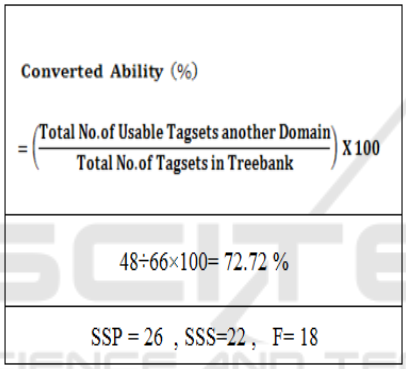

Figure 2: Check converted ability Formula

It means that URDU.KON-TB is already converted in dependency parsing domain. So we just need to increase the feature according to dependency grammar rules and increase the accuracy. So that's why we don't need to develop new dependency Treebank. Another mean of this percentage is that, 72.72% words as a HEAD working in UEDU.KON-TB. So, dependency relationship not enhance according to this nature in Treebank.

In 2017, Munir et al, get minimum 49% accuracy because Functional tagset is not marked by dependency grammar rules. If we, marked few Functional tagset according to dependency grammar rules with assumption is every word in a sentence is Head value give zero show in figure 1 then must accuracy will be increased.

4 ENHANCE FEATURES

We are talk about MaltParser is popular for its dependency structure parsing is the set of rules used for describing asymmetric dependencies between a head and dependent adopted in figure 3.

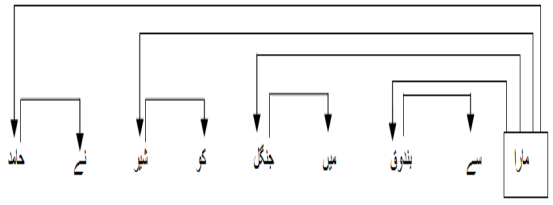

Figure 3: URDU.KON-TB Dependency Structure

We evaluated during the manual process of adding features in URDU.KON-TB. The word order in URDU.KON-TB is Sub+Obj1+ Obj2+Verb1 to Verb11.After enhance features, the precedence order is POS > Syntactic > Semantic & Functional tagset have grammatical information (sub, obj1, obj2, obl, plink and modf) and Semantic is relation between words (Munir et al., 2017) and (Ali et al., 2010).

The functional tagset marked according to rule of dependency grammar, which is every token contain three information. We consider the most frequent information of a token is used in URDU.KON-TB. After enhance features, we able to say that in 2017, Munir et al consider functional tagset is dependency relation as DEPREL is not against the dependency grammar rule just missing head information in Treebank (Munir et al., 2017). So, we just needed adding head information to explain dependency relation show in figure 3 and 4.

ID	FORM	-	POSTAG	CPOSTAG	-	HEAD	DEPREL	-	-
1	حامد	-	N.PROP	KP.ERG	-	9	KP.ERG.SUB	-	-
2	نے	-	CM	KP.ERG	-	1	KP.ERG.SUB	-	-
3	شیر	-	N	KP.ACC	-	9	KP.ACC.OBJ	-	-
4	کو	-	CM	KP.ACC	-	3	KP.ACC.OBJ	-	-
5	جنگل	-	N.SPT	KP.SPT	-	9	KP.SPT.MODF	-	-
6	میں	-	CM	KP.SPT	-	5	KP.SPT.MODF	-	-
7	بندوق	-	N	KP.INST	-	9	KP.INST.MODF	-	-
8	سے	-	CM	KP.INST	-	7	KP.INST.MODF	-	-
9	مارا	-	V.PERF	VCMAIN	-	0	VCMAIN.Root	-	-

Figure 4: CoNLL Format

MaltParser based on Data-driven dependency parsing Approach in which we map input strings to output. The CoNLL format data are given to MaltParser as input. It allows user-defined feature

models contain lexical, part-of-speech and dependency feature as ID, FORM, POSTAG, CPOSTAG, HEAD and DEPREL. MaltParser use Nivre algorithm to train and test data (Munir et al., 2017) ,(Ali et al., 2010) , (J.nivre, 2006) and (Abbas, 2014).

5 MODEL

Architecture and computational model is an Urdu Dependency Parsing System-2(DPS2) showing in figure 5. We have used proposed Data-Driven Dependency Parsing computational model (Munir et al., 2017) and (Ali et al., 2010).

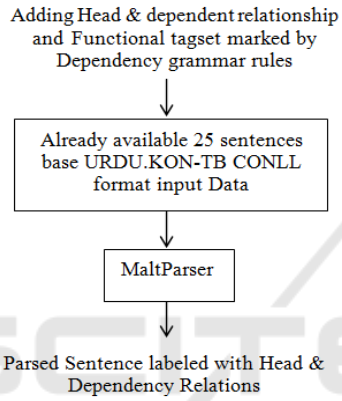

Figure 5: Model is Dependency Parsing System-2 (DPS2).

6 EXPERIMENTS

The experiment performed to check accuracy in dependency parsing system-2. The 25 sentences of URDU.KON.TB CoNLL format input data already available. Just need to enhance features. After that, we splitting it to 80% trained data and 20% tested data is given to MaltParser using Nivre arc-eager algorithm for parsing. In this way, 8 experiments are possible. But, we conduct only one experiment with six different feature models show in table 1.To check increase the features in URDU.KON-TB for increase the accuracy.

7 RESULT

The correctness of DEPREL tag is comparing MaltParser parse output with manually tagged test data. The accuracy percentage of experiment is calculated using this formula:

$$Accuracy\ (\%) = \left(\frac{Total\ No.of\ correct\ marked\ DEPREL}{Total\ No.of\ tokens}\right) X\ 100$$

The accuracy of 48/67*100=71.641 percentage is noted of this experiment show in table 1. We able to say, that URDU.KON-TB is the dependency structure base Treebank. The results also show increasing the features in the model is helpful to increase the accuracy to support our finding and argument. The comparison of accuracy also shows in figure 6 with (Munir 1 et al., 2017).

In future work, we adding boundary of phrases in URDU.KON-TB, automatically give mini 500 sentences to MaltParser that's why, we need to tune MaltParser and claim final accuracy. In this way, we can conduct more experiment and finally reported usefulness, errors and issues as proposed by Ali and Hussain (Ali 1 et al., 2010).

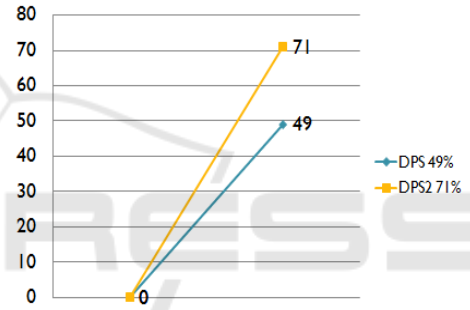

Figure 6: Comparison of Accuracy

Table 1: Result.

No.	Experiments with Feature Model	Accuracy (%)
1	ID, FORM, POSTAG (SSP), CPOSTAG (SSS), HEAD, DEPREL	**71.641**

REFERENCES

Munir, S., Abbas, Q., & Jamil, B. "Dependency Parsing using the URDU. KON-TB Treebank". *International Journal of Computer Applications* 2017, *167*(12).

Abbas, Q. (2014). Building Computational Resources: The URDU. KON-TB Treebank and the Urdu Parser (Doctoral dissertation).

Ali, W., &Hussain, S. (2010). Urdu dependency parser: a data-driven approach. In Proceedings of Conference

on Language and Technology (CLT10), SNLP, Lahore, Pakistan.

Ali, W, (2010). Data-Driven Dependency Parsing for Urdu, MS (MPhil), Computer Sciences thesis, Department of Computer Sciences, National University of Computer and Emerging (NUCES), Lahore, Pakistan.

J. Nivre, Inductive Dependency Parsing, Springer, 2006.

M. Marcus, B. Santorini, and M.A. Marcinkiewicz, "Building a large annotated corpus of English: The Penn Treebank", Computational Linguistics 1993.

Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., ... &Marsi, E. (2007). MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, *13*(02), 95-135.