Assignment 2 Report

Yashar Kor
University of Alberta
CCID: yashar
yashar@ualberta.ca

1 Definition of Performance Metrics

There are various performance measures that represent the correctness of classification. One of the most popular sets of such measures includes: precision, recall, and F_1 score. In order to provide their definitions, a *confusion matrix* is introduced where **UnGr** means Ungrammatical, and **Gr** Grammatical:

		Real Value	
		UnGr	**Gr**
Pred	**UnGr**	$TruePositive : TP$	$FalsePositive : FP$
	Gr	$FalseNegative : FN$	$TrueNegative : TN$

False positive and false negatives are defined as follows. FP (false positive): the sentence is grammatical but your checker found it not. FN (false negative): the sentence is ungrammatical but your checker labelled it as 'grammatical'. The performance measures are defined in the following way:

$$precision = \frac{TP}{TP + FP} \quad recall = \frac{TP}{TP + FN}$$

and

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

2 Evaluation

In this section the performance of the grammar checker is evaluated. In table (1) the precision, recall, and f1-score are reported. The grammar checker has 30 % precision in finding the ungrammatical sentences which means out of all predicted 'ungrammatical', 30% were correctly classified. Furthermore, the grammar checker has 68% recall which means out of all ungrammatical sentences, the developed model could detect 68 % of them.

The grammar checker produced **false positive** since:

- the grammar rules developed in toy.cfg file does not cover all grammatical sentences and has limited rules, to find correct grammatical sentences. Thus, there are some correct grammatical sentences that can not be identified by our developed grammar in toy.cfg file which also decrease precision.
- In this assignment instead of parsing the actual sentences we wrote grammar to parse the POS tag generated by computer models. Therefore, POS tags can have error which can produce both false positive or false negative

The grammar checker produced **false negative** since:

- grammar rules in toy.cfg are too general which means cannot detect the nuance grammatical difference of sentences. Therefore, it can detect ungrammatical sentence as gramatically correct which add FN samples.
- some verbs or nouns behave differently and need specific grammar for themselves. In this assignment we did not consider the various verbs, or various nouns, since we removed the lexicons. They need different grammatical structure, and ignoring them can add to the FN samples.
- We implemented the grammar checker on POS tags which means we removed semantics from sentences. Semantics can change the meaning and a sentence which seems grammatically correct from POS point of view, can be labelled ungrammatical.

3 Classification report

In this section the more detailed classification report table (2) for both lables 'Grammatical' which

precision	recall	f1-score
0.30	0.68	0.41

Table 1: Classification Report

means grammatical sentences and label 'Ungrammatical' which means ungrammatical sentences are provided.

	precision	recall	f1-score	support
Grammatical	0.69	0.31	0.43	423
Ungrammatical	0.30	0.68	0.41	180
accuracy			0.42	603
macro avg	0.50	0.49	0.42	603
weighted avg	0.58	0.42	0.42	603

Table 2: Classification Report

References