NLPCC 2018 Shared Task Guideline:

Grammatical Error Correction

1. Task Definition

Automatically correcting grammatical errors is a challenging task which has attracted increasing attention recently. Our goal is to detect and correct grammatical errors present in Chinese essays written by non-native speakers of Mandarin Chinese. Given annotated training data with corrections of grammatical errors and blind test data, the participating teams are expected to submit automatically corrected version of texts in test data.

2. Data

The training data was collected from Lang-8[1], a language-learning website where native speakers freely choose learners' essays to correct. Noting that some input sentences may have different corrections. The test data was texts written by foreign students and carefully corrected by professors. Both datasets was collated into the same form. Samples are shown in Table 1.

Table 1   Samples from training data.

Original Sentence	Corrected Sentences
长成大人，我盒饭做的很开心。	长大成人后，我做盒饭做得很开心。
城市里的人能度过多方面的生活。	城市里的人能过丰富多彩的生活。
	城市里的人能过多种多样的生活。
	城市里的人能过多方面的生活。

3. Evaluation

We use M^2 Scorer to evaluate performances.

The precision, recall, and F_1 measure between the set of system edits $\{e_1, \ldots, e_n\}$ and the set of gold edits $\{g_1, \ldots, g_n\}$ for all sentences are computed as following formulas:

$$P = \frac{\sum_{i=1}^{n} |e_i \cap g_i|}{\sum_{i=1}^{n} |e_i|}$$

$$R = \frac{\sum_{i=1}^{n} |e_i \cap g_i|}{\sum_{i=1}^{n} |g_i|}$$

$$F_1 = 2 \times \frac{P \times R}{P + R}$$

where the intersection between e_i and g_i is defined as:

$$e_i \cap g_i = \{e \in e_i | \exists g \in g_i (match(e, g))$$

Here is an example.

Original sentence: 随着通迅技术的发达我们的生活也是越来越放便。

Gold-standard edit set: g={通迅 → 通讯, 也 → ∈, 放便 → 方便} (∈ denotes empty character)

System edit set: e={通迅 → 通讯, 放便 → 方便}

According to above formulas, we can calculate P, R and F_1: P=2/3, R=2/2=1, F_1=2RP/(R+P)=4/5.

MaxMatch (M^2) computes the sequence of phrase-level edits between a source sentence and a system hypothesis that achieves the maximal overlap with the gold standard.