Classification of Rating & Tags for Codeforces Competitve Programming Problems

Nilay Tripathi   Nicholas Belov   Sriram Ramakrishnan

October 20, 2024

1 Background & Introduction

Codeforces[1] is a website centered around competitive programming, where problems are designed to test ingenuity in mathematics and coding knowledge. Codeforces allows users to participate in contests, which consist of several problems. Each problem is assigned a numeric rating which correlates with the problem's general level of difficulty. Each problem is also assigned specific tags, which give the mathematical and/or programming skills that are deemed most relevant to solving the problem.

For our project, we would like to construct a model which predicts either the numeric difficulty rating or the tags (or both, if time and resources permit).

2 Obtaining Data

We think this project neatly fits into the "novel dataset" category. Obtaining data will likely be achieved through scraping the Codeforces website with a Python scraping script to retrieve the problem text, tags, and rating for a subset of the problems on the website.

3 Methods

The most appropriate model for this task seems to be a recurrent neural network (RNN), but we may also explore more complicated models such as transformers to extract features and patterns from the source text.

---

[1] https://codeforces.com/