Focus on coding style: Only consider working predictions. Predict the competitive programming provides an opportunity for... 

For bigrams/sequence the lowest accuracy is 69.6% on test and 77.2% on train. Results (10-fold cross validation)

- Training on 9 contests (~54k examples) and testing on 1 contest (~6k examples) to test generalization to unseen problems.
- Accuracy is the weighted accuracy where the weight of each example is the inverse of the class size. This shows how well the model can predict all classes. (Unweighted accuracies are generally higher.)

Models

Linear regression (for rating)
- Loss for single example: \(-w \cdot \text{p}(x | y = k) \sim N(\mu, \sigma^2)\)
- Dropout for hidden nodes in training.

Gaussian discriminant analysis (GDA)
- \(p(y = k | x) = \frac{1}{N}(\mu_k, \Sigma_k)\)
- Gaussian distribution assumed for each class.

Logistic regression
- For country prediction, standard weighted softmax.
- For rank prediction, train 10 independent logistic models. Each example is part of ranks \(r \times 1, r \times 2\).
- Weights are inverse of class size.

Neural network
- Output layer and weighted loss functions are the same as logistic regression.
- 50% dropout for hidden nodes in training.

Preprocessing

Source code
```
int n = 0; var int main() { 
  scanf("%d", &n); 
}
```

libclang C++ parser

Codeforces Rank

<table>
<thead>
<tr>
<th>Codeforces Rank</th>
<th>Rating Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legendary Grandmaster</td>
<td>3000+</td>
</tr>
<tr>
<td>International Grandmaster</td>
<td>2600-2999</td>
</tr>
<tr>
<td>Grandmaster</td>
<td>2400-2599</td>
</tr>
<tr>
<td>International Master</td>
<td>2300-2399</td>
</tr>
<tr>
<td>Master</td>
<td>2100-2299</td>
</tr>
</tbody>
</table>

Data Set

- 10 Codeforces contests, Aug-Nov 2018.
- Selected countries are open to all ratings.
- Scraped with custom scraper.
- Only consider last passing submission. Only consider C++ (~70% of total).
- ~6k submissions per contest, total ~60k.
- For country analysis, only participants in the top 10 countries (~70% of total).

For International Grandmaster vs. Pupil, here are the features where the class means differ the most.

Strongest indicators of high skill level
- Use of if/elif, assert, and function templates.
- Use of cin/cout instead of scanf/printf, or perhaps cin >> x; cin >> y; instead of cin >> x >> y;
- It seems that the model uses some features to favor shorter solutions. TranslationUnit is always present exactly once, so the normalized value will be higher in shorter solutions.

Strongest indicators of low skill level
- Use of cin/cout instead of scanf/printf, or perhaps cin >> x; cin >> y; instead of cin >> x >> y;
- Use of #ifdef, assert, and function templates.

References

- Koh, Jun, Sreejith; Lee and Tan; and Sui; they were used to implement this project.

Future Work

- More data will likely help. Going from 5 to 10 contests increased the accuracy significantly.
- Try a recurrent neural network, e.g. LSTM.
- Improve token processing, e.g. also replace class and macro names with placeholders.
- Interpretation of the neural network model.