Every Team Deserves a Second Chance: Identifying when Things Go Wrong
(Student Abstract Version)

Vaishnavh Nagarajan¹, Leandro Soriano Marcolino², Milind Tambe²
¹ Indian Institute of Technology Madras, Chennai, Tamil Nadu, 600036, India, +91-44-22574350
http://www.cse.iitm.ac.in/~vaish/
vaish@cse.iitm.ac.in
² University of Southern California, Los Angeles, CA, 90089, USA
{sorianom,tambe}@usc.edu

Appendix
Reduced Vector of Features
As mentioned, we also study a variant of our prediction method, where we use only information about the number of agents that agreed upon the chosen action, but not which agents exactly were involved in the agreement. For that variant, we consider a reduced feature vector \( \vec{y} = (y_0, y_1, \ldots) \), where we define \( y_i \) to be the proportion of times that the chosen action was agreed upon by any subset of \( i \) agents:

\[
y_i = \frac{\sum_{k=0}^{||M||-1} \mathbb{I}(||H_k|| = i)}{||M||},
\]

where \( \mathbb{I} \) is the indicator function and \( M_j \subseteq M \) is the set of world states from \( m_0 \) to the current world state \( m_j \).

In order to show that we can also have high quality predictions with a much more scalable representation, we also run experiments using the reduced feature vector for all 3 teams. In Figure 3 we can see the results when predicting in the end of the games with the reduced feature vector. When comparing the results with the full representation, we notice that the accuracy does not change much for diverse and intermediate, and the difference is not significant (\( p = 0.9929 \) and \( p = 0.8403 \), respectively). For uniform we observe an improvement in the accuracy of 4%, but such improvement is also not statistically significant (\( p = 0.2867 \)).

In Figure 4 we can see the prediction at each stage of the game, again comparing with Perfect's evaluation. As we can see, we can also obtain a high accuracy quickly with the reduced feature vector, reaching 60% again towards the middle of the games. This time, there is less difference in the accuracy for the diverse and uniform teams, but we can still show that the accuracy for diverse is significantly better than for uniform (with \( p < 0.1 \)) in 15% of the stages (20% including a stage where \( p \approx 0.1 \)). Note that, again, the accuracy for the intermediate team is close to the one for uniform, even though intermediate is a significantly weaker team.
Figure 4: Performance metrics over all turns of 691 games, using the reduced feature vector.