Identify Unfriendly Robots

<u>Problem Definition</u>

A cyber police received the following information about Unfriendly and Friendly Robots:
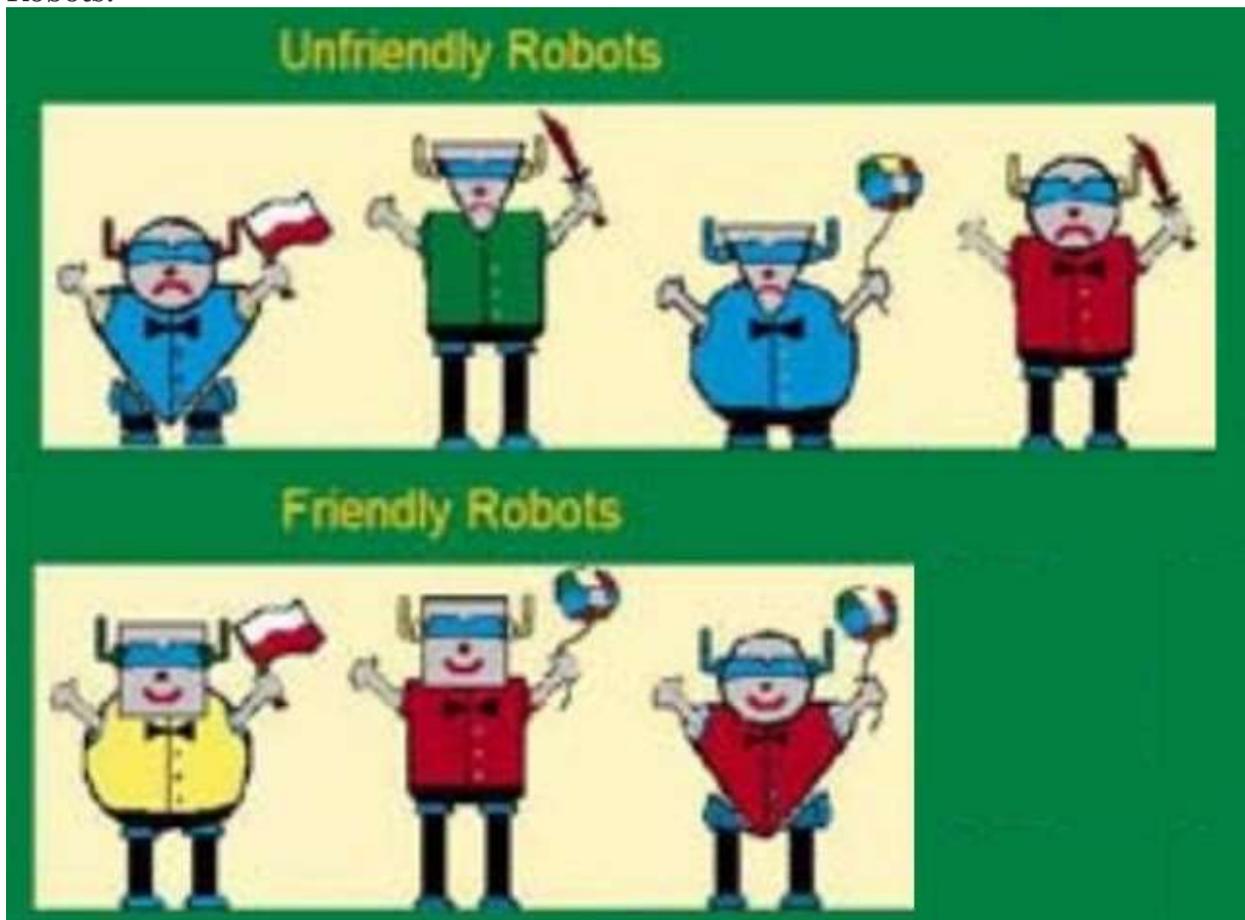
The police asked its analysts to specify rules to identify if a robot is friendly or unfriendly, The rules were expressed in terms of any features you can see in robots, such as the shape of the head, the color of the jacket, the height, the color of their antennas, what they are holding in their hands, whether they are smiling or not, etc. The analysts quickly identified simple rules that succeeded for the robots from the above lists. Can you guess which rules were used?

However, when the police received more information about robots the manually determined rules failed miserably.
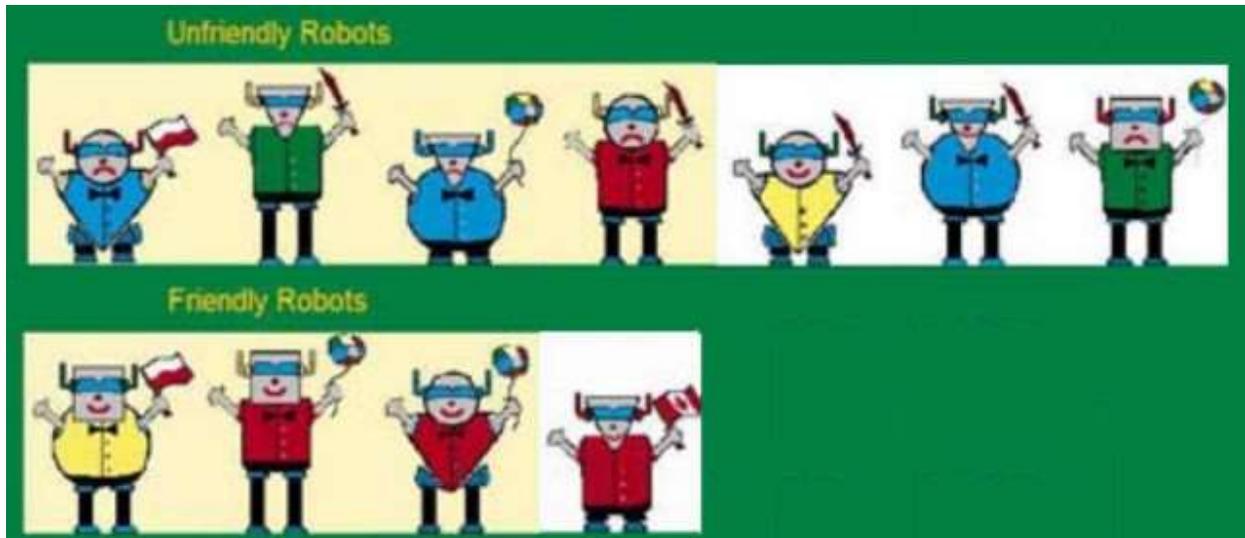
Please help the police to determine rules that can identify if any robot (from these lists or a new one) is friendly or unfriendly. You may use any combination of machine learning and business rules tools to come up with reliable rules.

The driving principles for provided below solutions are the logic and simplicity

Solution 1

Based on initial info, received by police, it doesn't take too long to conclude that all unfriendly robots have an angry face. Just such a single "bad feature" distinguishes them from friendly robots, all of which have quite opposite a smily face. In other words our logic for creating rules, identifying unfriendly robots, is quite simple - such rule should be Not True for any of friendly robots. This logic allows to create Rule 1.1 (the first Rule for Solution 1) to identify unfriendly robots:

Rule 1.1: robot has angry face

Advantage of Rule 1.1 its self-explanatory, i.e. all unfriendly robots are angry, and quite opposite all friendly robots are smily.
Appearance of additional info, obtained by police, shows that Rule 1.1 is not enough for identifying unfriendly robots, it should be added by Rule 1.2

Rule 1.2: robot is holding a sword

Again advantage of Rule 1.2 has same advantage as Rule 1.1 - it explains itself and easy to understand, i.e. a sword in the hands of robot is enough to consider such robot unfriendly. This allows to say that robot is unfriendly if either Rule 1.2 OR Rule 1.1 is

valid. In other words we logically derived combined Rule for identifying unfriendly robots in Solution1:

     Solution 1: robot is holding a sword OR robot has angry face.

Opponents of Solution 1 may say that Rule 1.1 turns out to be deceiving, since some

unfriendly robots (like the one in yellow jacket - _____) may smile and pretend to be friendly, while holding a sword. To counter such argument, we can say that Solution 1 evaluates Rule 1.2 first. So, robot in yellow jacket and holding a sword will be immediately identified as unfriendly, although having a smily face. As alternative approach it's possible to completely exclude the presence of smile from our solution. This leads to the following Solution 2:

Solution 2

Deriving from the knowledge, obtained from analyzing Solution 1, we can say that "holding a sword" may be the main indication of unfriendly robots. Besides Based on initial info, received by police, all other unfriendly robots wear a blue jacket. Again, for creating our rules we are considering the same logic as in Solution 1 - our rules should be Not True for any of friendly robots. This logic allows creating two initial rules for identifying unfriendly robots:

     Rule 2.1: robot is holding a sword
     Rule 2.2: robot has a blue jacket

Appearance of additional info, obtained by police, shows that Rule 2.3 needs to be added for identifying unfriendly robots

     Rule 2.3: robot has a green jacket

All above allows to logically derive combined Rule for identifying unfriendly robots in Solution 2:

     Solution 2: robot is holding a sword OR robot has a blue jacket OR robot has a green jacket

Opponents of Solution 2 may say it is less self-explanatory and more complex than Solution 1. Although these arguments are valid we can mention advantage of Solution 2 - it eliminates the presence of angry face from the rules set, since based on Solution 1 analysis angry face may be deceiving rule in identifying unfriendly robots.

Thx, Vlad, vsilverman@gmail.com
https://github.com/vsilverman

P.S. Please vote which of above solutions you like better and if possible, explain why

P.P.S. Providing alternative solutions would be always appreciated.