FIT3080 Assignment 3

Lawrence MacDonald, Dekel Pilli
{lmmac8,dpil7}@student.monash.edu

October 21, 2016

1 Soccer

a) $V^\pi(1) = \max_S \sum_{s'} T(1,S,s')[R(1,S,s') + \gamma V^\pi(s')]$
$V^\pi(1) = \frac{1}{6}[1 + 1 \times 0] + \frac{5}{6}[0 + 1 \times 0]$
$V^\pi(1) = \frac{1}{6} + 0$
$V^\pi(1) = \frac{1}{6}$

b) $Q^*(3,D) = \sum_{4} T(3,D,4)[R(3,D,4) + \gamma V^*(4)]$
$V^*(4) = \frac{2}{3}$, using the same logic as part a), as the only possible action from state 4 is S.
$Q^*(3,D) = y[0 + \frac{2}{3}]$
$Q^*(3,D) = \frac{2y}{3}$

c) $V_0(1) = 0$
$V_0(2) = 0$
$V_0(3) = 0$
$V_0(4) = 0$
$V_0(G) = 0$
$V_0(M) = 0$
Because G and M are terminal states, their value will not change across iterations.

Because all non-terminal states in iteration 0 have value 0 and reward 0, any D actions can be assumed to return a value of 0 in iteration 1:
$V_1(s,D) = \frac{x}{6}[0 + 1 \times 0] + \frac{1-x}{6}[0 + 1 \times 0]$
Hence, all values in iteration 1 will be calculated with action S.
$V_1(1) = \sum_{s'} T(1,S,s')[R(1,S,s') + \gamma V_0(s')]$
$V_1(1) = \frac{1}{6}[1 + 1 \times 0] + \frac{5}{6}[0 + 1 \times 0]$
$V_1(1) = \frac{1}{6}$
$V_1(2) = \frac{1}{3}[1 + 1 \times 0] + \frac{2}{3}[0 + 1 \times 0]$
$V_1(2) = \frac{1}{3}$
$V_1(3) = \frac{1}{2}[1 + 1 \times 0] + \frac{1}{2}[0 + 1 \times 0]$

$V_1(3) = \frac{1}{2}$
$V_1(4) = \frac{2}{3}[1 + 1 \times 0] + \frac{1}{3}[0 + 1 \times 0]$
$V_1(4) = \frac{2}{3}$

Values in any iteration for action S will never change, as the only next states, and hence variables that the value iteration algorithm relies on, are terminal states and thus never change in value or reward.
$V_2(1, S) = \frac{1}{6}$
$V_2(1, D) = \frac{3}{4}[0 + 1 \times \frac{1}{3}] + \frac{1}{4}[0 + 1 \times 0]$
$V_2(1, D) = \frac{1}{2}$
Hence, $V_2(1) = \frac{1}{2}$
$V_2(2, S) = \frac{1}{3}$
$V_2(2, D) = \frac{3}{4}[0 + 1 \times \frac{1}{2}] + \frac{1}{4}[0 + 1 \times 0]$
$V_2(2, D) = \frac{3}{8}$
Hence, $V_2(2) = \frac{3}{8}$
$V_2(3, S) = \frac{1}{2}$
$V_2(3, D) = \frac{3}{4}[0 + 1 \times \frac{2}{3}] + \frac{1}{4}[0 + 1 \times 0]$
$V_2(3, D) = \frac{1}{2}$
Hence, $V_2(3) = \frac{1}{2}$
Because S is the only possible action from state 4, and values do not change across iterations for action S (because the two next states are both terminal), $V_2(4) = \frac{2}{3}$

d) $V_3(1, S) = \frac{1}{6}$
$V_3(1, D) = \frac{3}{4}[0 + 1 \times \frac{3}{8}] + \frac{1}{4}[0 + 1 \times 0]$
$V_3(1, D) = \frac{9}{32}$
Hence, $V_3(1) = \frac{9}{32}$
$V_3(2, S) = \frac{1}{3}$
$V_3(2, D) = \frac{3}{4}[0 + 1 \times \frac{1}{2}] + \frac{1}{4}[0 + 1 \times 0]$
$V_3(2, D) = \frac{3}{8}$
Hence, $V_3(2) = \frac{3}{8}$. Because this is equal to the second iteration's value, state 2's value has converged.
$V_3(3, S) = \frac{1}{2}$
$V_3(3, D) = \frac{3}{4}[0 + 1 \times \frac{2}{3}] + \frac{1}{4}[0 + 1 \times 0]$
$V_3(3, D) = \frac{1}{2}$
Hence, $V_3(3) = \frac{1}{2}$. Because this is equal to the second iteration's value, state 3's value has converged.
Because, as described in part c, state 4's value does not change across iterations, its value has converged.

$V_4(1, S) = \frac{1}{6}$
$V_4(1, D) = \frac{3}{4}[0 + 1 \times \frac{3}{8}] + \frac{1}{4}[0 + 1 \times 0]$
$V_4(1, D) = \frac{9}{32}$
Hence, $V_4(1) = \frac{9}{32}$. Because this is equal to the second iteration's value, state 3's value has converged.
Thus, after four iterations, all states' values have converged.

e) From part b, $Q^*(3, D) = \frac{2y}{3}$
$Q^*(3, S) = \frac{1}{2}[1 + 1 \times 0]$
$Q^*(3, S) = \frac{1}{2}$
Thus, $\frac{2y}{3} \leq \frac{1}{2}$
$2y \leq \frac{3}{2}$
$y \leq \frac{3}{4}$
Hence, for $Q^*(3, S) \geq Q^*(3, D)$, $0 \leq y \leq \frac{3}{4}$

f)

$$R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & G & M \end{array} \\ \begin{pmatrix} & 0 & 0 & 0 & 1 & 0 \\ & & 0 & 0 & 1 & 0 \\ & & & 0 & 1 & 0 \\ & & & & 1 & 0 \end{pmatrix} \end{array}$$

$$initQ = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cc} D & S \end{array} \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ & 0 \end{pmatrix} \end{array}$$

If any samples of (s, a, s') return 0 and (s, a) is currently 0 on the Q-matrix, a new Q value will not be calculated, as it will remain 0.

Episode: $1 - D, 2 - D, 3 - D, 4 - S, G$
$sample(1, D, 2) = R(1, D, 2) + \gamma \max_{a'} Q(2, a')$
$sample(1, D, 2) = 0 + 1 \times 0$
$sample(1, D, 2) = 0$
$sample(2, D, 3) = 0 + 1 \times 0$
$sample(2, D, 3) = 0$
$sample(3, D, 4) = 0 + 1 \times 0$
$sample(3, D, 4) = 0$
$sample(4, S, G) = 1 + 1 \times 0$
$sample(4, S, G) = 1$
$Q(4, S) = (1 - \frac{1}{2})Q(4, S) + \frac{1}{2} sample$
$Q(4, S) = \frac{1}{2} \times 0 + \frac{1}{2} \times 1$
$Q(4, S) = \frac{1}{2}$

$$newQ = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cc} D & S \end{array} \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ & \frac{1}{2} \end{pmatrix} \end{array}$$

Episode: $1 - D, 2 - D, 3 - D, 4 - S, M$
$sample(1, D, 2) = 0 + 1 \times 0$

$sample(1, D, 2) = 0$
$sample(2, D, 3) = 0 + 1 \times 0$
$sample(2, D, 3) = 0$
$sample(3, D, 4) = 0 + 1 \times \frac{1}{2}$
$sample(3, D, 4) = \frac{1}{2}$
$Q(3, D) = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{1}{2}$
$Q(3, D) = \frac{1}{4}$
$sample(4, S, M) = 0 + 1 \times 0$
$sample(4, S, M) = 0$
$Q(4, S) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 0$
$Q(4, S) = \frac{1}{4}$

$$newQ = \begin{array}{c c} & \begin{array}{cc} D & S \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{4} & 0 \\ & \frac{1}{4} \end{pmatrix} \end{array}$$

Episode: $1 - D, 2 - D, 3 - S, G$
$sample(1, D, 2) = 0 + 1 \times 0$
$sample(1, D, 2) = 0$
$sample(2, D, 3) = 0 + 1 \times \frac{1}{4}$
$sample(2, D, 3) = \frac{1}{4}$
$Q(2, D) = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{1}{4}$
$Q(2, D) = \frac{1}{8}$
$sample(3, S, G) = 1 + 1 \times 0$
$sample(3, S, G) = 1$
$Q(3, S) = \frac{1}{2} \times 0 + \frac{1}{2} \times 1$
$Q(3, S) = \frac{1}{2}$

$$newQ = \begin{array}{c c} & \begin{array}{cc} D & S \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 0 \\ \frac{1}{8} & 0 \\ \frac{1}{4} & \frac{1}{2} \\ & \frac{1}{4} \end{pmatrix} \end{array}$$

Episode: $1 - D, 2 - S, M$
$sample(1, D, 2) = 0 + 1 \times \frac{1}{8}$
$sample(1, D, 2) = \frac{1}{8}$
$Q(1, D) = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{1}{8}$
$Q(1, D) = \frac{1}{16}$
$sample(2, S, M) = 0 + 1 \times 0$
$sample(2, S, M) = 0$

$$newQ = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cc} D & S \end{array} \\ \begin{pmatrix} \frac{1}{16} & 0 \\ \frac{1}{8} & 0 \\ \frac{1}{4} & \frac{1}{2} \\ & \frac{1}{4} \end{pmatrix} \end{array}$$

Episode: $1-D, 2-D, 3-D, M$

$sample(1, D, 2) = 0 + 1 \times \frac{1}{8}$

$sample(1, D, 2) = \frac{1}{8}$

$Q(1, D) = \frac{1}{2} \times \frac{1}{16} + \frac{1}{2} \times \frac{1}{8}$

$Q(1, D) = \frac{3}{32}$

$sample(2, D, 3) = 0 + 1 \times \frac{1}{2}$

$sample(2, D, 3) = \frac{1}{2}$

$Q(2, D) = \frac{1}{2} \times \frac{1}{8} + \frac{1}{2} \times \frac{1}{2}$

$Q(2, D) = \frac{5}{16}$

$sample(3, D, M) = 0 + 1 \times 0$

$sample(3, D, M) = 0$

$Q(3, D) = \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times 0$

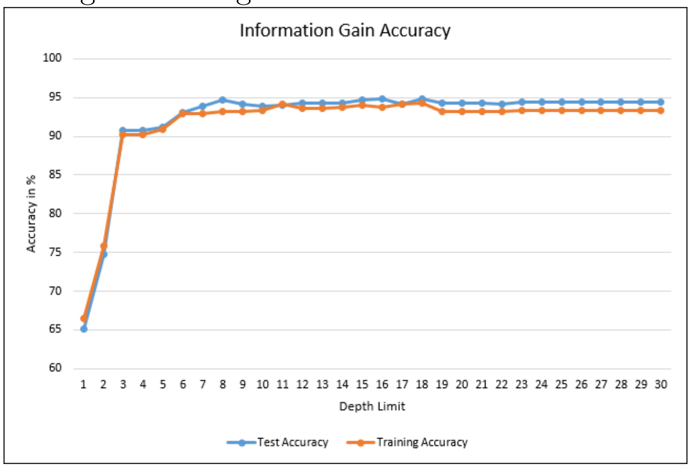
$Q(3, D) = \frac{1}{8}$

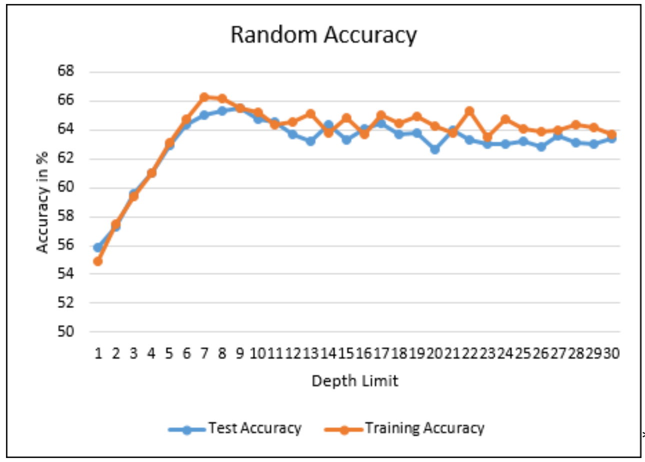
$$finalQ = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cc} D & S \end{array} \\ \begin{pmatrix} \frac{3}{32} & 0 \\ \frac{5}{16} & 0 \\ \frac{1}{8} & \frac{1}{2} \\ & \frac{1}{4} \end{pmatrix} \end{array}$$

No more episodes, so finalQ is the final result of our Q-learning from this data set.

2 Decision Trees

When using Information Gain to collect data, there was a clear case of diminishing returns on an increase in depth vs. accuracy. Certain thresholds even showed worse accuracy than some lower values, such as 9 depth, which was around 1% less accurate than 8 depth for both the test and training accuracy.
Similarly, when using the random splitting method, the accuracy decreased slightly in the region of 8-10 depth limit. Diminishing returns on increasing the depth limit were also seen with Random, with a depth limit of 200 providing minimal to no increase in accuracy for testing or training data.





*

As can be seen in the graphs, the testing data accuracy and training data accuracy was very close for both splitting methods, with Information Gain providing considerably better results than Random.
Information gain and depth limit of 8 gave the best accuracy for the testing data, but all information gain limits above 5 gave results very close to 94%. At depth limits lower than 5, not all of the attributes can fit on the tree, which limits the potential accuracy. This can be seen in the Information Gain Accuracy graph, where the values change very little from 6 onward.
Information gain, paired with the highest possible depth limit, gave the best accuracy for the training data. Again, from a bound of 6 or more, the results were very similar (within 0.5% of 93%), for the same reason mentioned above.
The largest differences between training data and testing data came at depth limit 22 (1.96%) using random. This large variance can be explained by the random nature of attribute selection for the nodes of the decision tree. Additionally, the tree has reached a level of information where it can predict itself well, because the random data is based on its own data, whereas that random allocation will have less correlation to the test data, though with more random tree learning, it begins to over-predict itself and achieve worse results.
*Random accuracy may vary. Collected as an average of 101 results.

3 Appendix A: Peer Assessment

We worked well together, utilising an Overleaf Latex document to work on the report, and a subversion repository to share the relevant files for both the report and the code. Apart from some very simple methods and questions which were implemented separately, both members contributed to reaching the answers. Complicated methods in the code that were implemented separately were later refined and bug-checked by the other team member. Furthermore, both members have ensured that they understand all solutions provided and how they were found. Overall, the workload was evenly split.