Explanation of method we used

We used an algorithm called "value iteration", in which for every node we give the value of the best Qvalue(reward+value of next state)). We continuously update the nodes until none of them change(and thus each node reached its optimal value). For nodes that are near "unknown" edges, we set its value as the node which is identical to the original but differs in that the edge is "open" multiplied by the probability it is open, plus the "closed" node multiplied by the probability it is closed.