CS 271 Spring 2020
Assignment 0

Student ID: 014498887
Name: Yash Sahasrabuddhe

Q1. Briefly (1 paragraph) summarize how an HMM is trained.

To train HMM, we know that the size of the matrices remains the same as we are trying to determine the elements of the matrices. That is, we have to find the parameters which best fit the observations.

This can be done keeping in mind that the matrices hold the row stochastic condition.

Using the gammas and the di-gammas, we can re-estimate the model.
We write the di-gamma function in terms of α, β, A, B matrices which are the functions generated from the back tracking and forward tracking algorithms.

As re-estimation is an iterative process, we need to have a stopping point. We define it as when the probability of P(O| λ) does not increase or some pre-defined number of iterations are completed.

According to this, we can summarize it in the following steps: -

1. Initialize the model, λ = (A, B, π).
2. Compute the $\alpha_t(i)$, $\beta_t(i)$, $\gamma_t(i)$ & $\gamma_t(i,j)$.
3. Based on these then, we can then re-estimate the model λ = (A, B, π) using the following equations: -

$$a_{ij} = \frac{\sum_{t=0}^{T-2} \gamma_t(i,j)}{\sum_{t=0}^{T-2} \gamma_t(i)} \quad \& \quad b_j(k) = \frac{\sum_{t \in \{0,1,\ldots,T-1\}} \gamma_t(j)}{\sum_{t=0}^{T-1} \gamma_t(j)}$$

   Where, a_{ij} is the transition probability from state q_i to state q_j and $b_j(k)$ is the probability of observing k in state q_j.

4. If P(O| λ) increases, we will again compute the functions and re-estimate until the stopping condition mentioned above are reached.

In this way, we can train an HMM model.

Q2. How is a trained HMM used to score a sequence?

To score a sequence, we need to find out the probability function P(O| λ) across a set of observations.

The most efficient way to find this is by using the forward pass or α-pass algorithm which is defined by

$$\alpha_t(i) = P(O_0, O_1, \ldots\ldots, O_t, x_t = q_i \mid \lambda)$$

Where $\alpha_t(i)$ is the probability of partial observation sequence up to time t, where underlying Markov process is in state q_i at time t.

Finally, P(O| λ) can be calculated as-

$$P(O \mid \lambda) = \sum_{i=0}^{N-1} \alpha_{T-1}(i)$$

The $\alpha_t(i)$ function can be calculated recursively which reduces the computation significantly i.e. from exponential time to pseudo-polynomial time.

This is how we can score a sequence.

Q3. Very briefly explain how an HMM and dynamic program differ.

DP focuses on finding the best probability at each state rather than finding the best path.

This probability is the probability of finding the best path ending at each of the states but not the path itself.

By keeping track of each state in the DP approach, we can find the optimal path by tracking back from highest scoring final state.

As DP approach only stores the highest scoring paths at each possible state and not the path itself, it makes the DP approach efficient.

The main difference from HMM is that in DP approach can be viewed as the α-pass algorithm but finding “max” instead of “sum” in the algorithm.

Q4. Why is it necessary to scale the values of the matrices when training an HMM?

When training the HMM, the values of the elements of the matrices and the functions like α,β etc. tend to 0 as the value of T increases in the functions and can cause underflows.

Hence, to avoid such underflows, it is imminent to scale the values of the numbers. This scaling though must be done diligently as the re-estimation formulae should remain valid post the scaling.

References:

[1] A Revealing Introduction to Hidden Markov Models by Mark Stamp (2018).

“Models will run the world”

a. In one paragraph, summarize the authors' main points.

The author has discussed the advantages as well as some of the disadvantages of the growing popularity of model driven business than the traditional data driven businesses. Not only the traditional software companies, but others such as Domino’s Pizza etc. have moved on to the model driven business strategies so that they can maintain their competitiveness and dominance in today’s market. The leaders in today’s industry such as Amazon, Netflix etc. have all focused-on learning models which are built on “closed loop” data. This means that once created, the model can learn from its own success and failure without any interruption of a programmed derived algorithms of a developer. The author has summarized that the data-driven model, data helps the business while in a model-driven business, the models are themselves the business. It then boils down to the industry who has the best resources and have the capacity to integrate the models into their businesses. The author then discusses some of the different industries who have changed their approach to a model driven approach. Furthermore, the author then discusses that businesses with less complete data sets have to rely on human intuition to develop and interpret the results. This limits their scope as well as advancement. There is an emphasis then that companies who want to become more model driven will need to have new disciplines of model management than the traditional ones. On the other hand, the most important disadvantage of using this model is that models themselves will have more issues. Hence, there comes the trust factor of whether you should rely on such results which may have a significant impact. In spite of this, companies who are hungry for such model driven advantage will grow in the coming near future.

b. Write second paragraph discussing what you most agree with and anything that you disagree with in this article.

The point with which I agree the most is that the models are self-learning, i.e. even if the result of some question is wrong, the model will still improve as the next time it will learn the true outcome and predict that. It doesn’t care about the failure or success of the test cases as it will always improve the model. Another point which can be strongly agreed upon is that in order for companies to maintain their dominance in the industry, it has to evolve with the new technologies and methods of businesses. One such example was given through the Domino’s Pizza company. Furthermore, it can be agreed that though these advancements will produce a lot of success, but with this also comes the negative aspects of the model driven business like how to keep the data private and prevent it from getting misused etc. It should be taken care of with great care. On the other hand, one can also argue if new disciplines are really required for maintaining this new model driven business. I think that this might be an area of debate as in case the companies trust is lost, this will incur a lot of liabilities.