Pattern Recognition and applications

Prof. P. K. Biswas
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 38
Hidden Markov Model

It the discussion on the problem of recognition of temporal patterns, that is we said that we can have patterns which actually unfold in time. So, the examples under this category are one you can consider is speech recognition problem, because we know that the speech signals are actually time baring signals. So, at different instance of time we have different signal properties or we have different signal strength, so speech is one of the signals which is actually time baring signal. For speech recognition purpose what people usually does is the speech the signal is divided into a number of phonemes.

You try to find out in which sequence the phonemes occur and based on that you try to recognize or you try to classify the word that has been spoken. So, the other applications in this temporal pattern recognition may be say activity recognition say for example, we want to find out whether some person is walking some person is running somebody is crawling and all that. It can be is a sign language recognition because sign language recognition is basically you expressed your ideas by using hand movements and it is the sequence in which the different parts of the hand move.

That actually keeps you or since of the information or you have to decode that particular sequence of hand gestures and from the sequence you have to interpret what message is been conveyed. Similarly, the activity recognition when I say running walking crawling this kind of applications is very useful for security surveillance purpose. So, we want to identify whether the movement of a person is a normal movement or it is an abnormal movement. So, through a video camera if I can capture that the movement of a person is that abnormal which obviously has to be detected based on the kind of movement a person does.

Obviously, it is a temporal sequence, so when I say the movement it is the body pose at different instance of time which occurs in a given sequence in a particular sequence and based on that the body poses that different instance of time. You try to identify whether the movement is a normal movement or abnormal movement and if the movement is abnormal you try to track that person because definitely a person going to an abnormal movement is a suspicious person. So, this temporal sequence or the patterns having temporality in it have lots of applications. I have just given two or three examples, but there are many other applications application domains in which this temporal sequence is need to be analyzed and they need to be classified.

Now, what we said in our previous class is that to recognize or to identify such temporal sequences, what we do is we make use of a machine which is something similar to with very similar to our sequential machine or finite state machine. So, as in we have a finite set of states through which the machine makes transition we have a finite set of output symbols we have a finite set of input symbols. So, if the machine is any state s at time instant t, the machine will go to another state at the next time instant depending upon the input that is fed and when it makes a transition to the next state it emits an output.

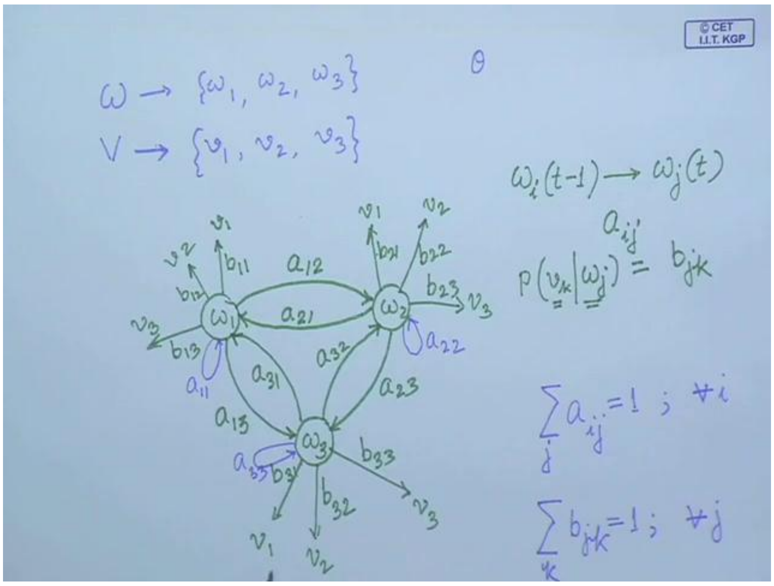
So, accordingly for a sequential machine I can have state transition table that is based on which input to which of the states the machine makes a transition. Similarly, I can also have a output table that for which type of transition for which type of transition from which state to which state what is the output the machine generates. It is this is output which actually absorbable state of a machine is not very absorbable.

So, I can identify the state from the input that is given and the output that is generated, so in this of a sequential machine we have said in this previous class in a sequential machine if the set of outputs is limited to only two symbols that is 0 and 1. The sequential machine of finite state machine actually becomes finite state automaton and such a finite state automaton has got huge application in sequence generation, beat sequence generation or beat sequence detection. So, beat sequence detection is a problem which is very useful in communication because this is what gives you the synchronization if I want to Identify a point from were see information starts.

So, that starting of the information can have a particular beat sequence and I can have a finite state automaton to detect the beat sequence and once the beat sequence is detected, then I can say that the next portion of the sequence. Actually, it gives the information, but that sequence which is detected is only for the synchronization purpose. So, similarly, in case of hidden Markov model, we have said that there are two types of states one kind of state is say omega.

So, this set of states we said that they are the hidden states these states cannot be observed by the perceiver and we have another type another type of a states which are actually visible states and that type of states is called say v. So, this is a state of visible states, so hidden Markov model can have a number of hidden states and it can have a number of visible states.

(Refer Slide Time: 07:29)



So, let us take an example having a hidden of model having say three states, so I have under omega I have states I have omega 1, omega 2 and omega 3. So, these are the three states of this hidden Markov model say we are saying a model as theta and v is the state of visible states or the symbols visible symbols which are emitted by this hidden Markov model when this HMM theta is in a particular state.

So, let us assume that this can generate say three visible symbols v 1 v 2 and v 3, so let us draw this HMM theta, so we have the states omega 1 omega 2 and omega 3. So, the machine of this hidden Markov model can make a transition from state omega 1 and omega 2 or this probability of transition is given by a 1 2. As we said that if the hidden Markov model is in state say omega i at time instance say p minus 1 and it Makes a transition to state omega j at time instant t, and the probability of such transition from state omega i to state omega g is demoted by a i j.

So, from omega one this hidden mark Markov model can make a transition to state omega two for the probability of transition is given by a 1 2, similarly omega 2 it can have a transition to 1 2 for the probability of transition is a 2 1. Similarly, a 1 3, a 1 2, it will be a 1 3 omega 3 to omega 1, this will be a p 1 omega 3 to omega 2 it will be a 3 2 omega 2 to omega 3 it will be a 2 3. So, these are transition probabilities between two different states, now in every state the machine can emit one of the visible states in each of these hidden states, the machine can emit one of the visible states.

So, from omega 1 it can generate it can emit the visible symbol or visible state omega v 1 with this probability of this emission which is given by v 1 1 because as we said that the probability that the machine emits v k when the machine is in state omega j. This actually given by b j k, so from state omega j it emits a symbol a visible symbol v k with the probability which is given by b j k. So, from omega 1 it emits a symbol v 2 with a probability v 1 2 it emits v 3 with probability b 1 3.

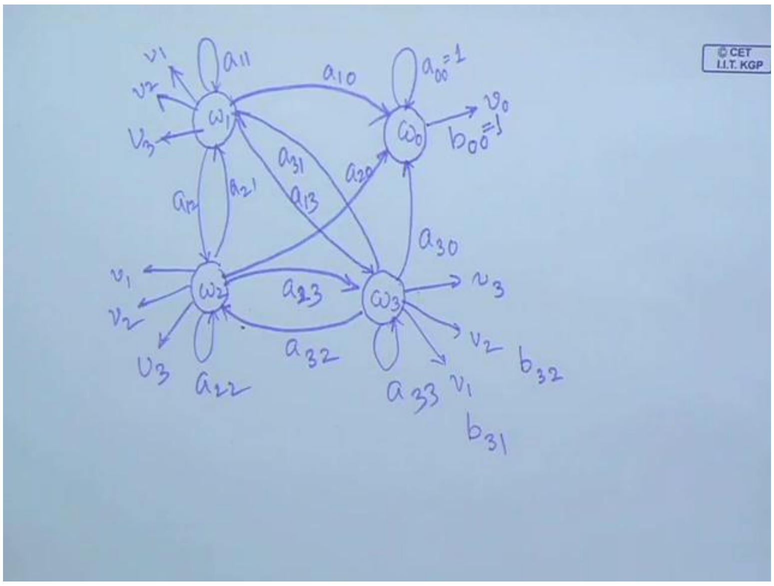
Similarly, it will be v 1 is emitted with probability a 2 1 v 2 is emitted with probability b 2 2 v 3 is emitted with b 2 3 from omega 3 v 1 is emitted with probability b 3 1 v 2 is emitted with probability b 3 2 and v 3 is emitted with b 3 3. So, these are the probabilities of emission of the visible states from different hidden states of this hidden Markov model. Now, find that given any state there will always be transition to some of the states including the state itself.

So, I can have a transition from omega 1 to omega 1 also and this transition probability will be a 1 1, similarly omega 2 to omega 2, this transition probability will be a 2 2. Similarly, omega 3 to omega 3, this transition probability will a 3 3, so this is also possible that is in two different time states to subsequent time states, the machine can be in the same state same hidden state. So, because from any hidden state there is always a transition to one of the states. So, naturally some of a i j for when I take the summation of a i j this will be equal to 1 and that will be true for all values of i because from any state the machine always makes a transition to one of the states.

So, the some of the probabilities a i j when I take the summation a i j that has to be equal to 1 and that has to be two for all and similarly when the machine is in any state it always emits a visible symbol. So, because it always emits the visible symbol, so I must have b j k when I take the summation over k, because v k is the symbol and b j a indicates that form hidden state omega j. The machine emits a visible symbol v k and it always emits a symbol so this b j k the sum of all k that also has to be equal to 1 and this will be true for all h that is for all the hidden notes.

So, given this and we have also said that hidden Markov model has a specific hidden state, which is called a receiving state or an accepting state or a final state and we have said that once the machine reaches the state it cannot come out of the state. So, all the transitions will be within the accepting the states only and while in that accepting state it will be emit one visible symbol. If you see it will not emit any other visible symbol, so by incorporating that accepting state into our hidden Markov model, I can redraw this hidden Markov model something like this.

(Refer Slide Time: 15:00)

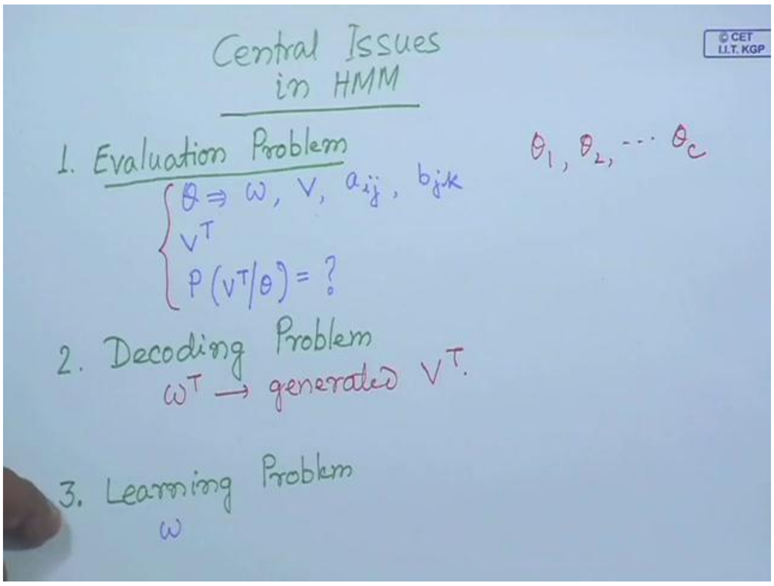


So, I have this omega 1, I have say omega 2, I have omega 3 and this is the accepting state which is omega 0, so from omega 1 to omega 2, I have transition a 1 2 omega 2 to omega 1 it is a 2 1. Similarly, all this different transitions, so after this I can have self transitions this is a 1 1, a 2 2, a 3 3, but from any of the state it can make a transition to omega 0. So, I can have this sort of transition, so here it will be here a 1 0, a 2 0, a 3 0, but once the hidden Markov model is in this final state omega 0 it cannot come back to any of the states. However, it can have transition within the state and because this all transition will all is omega 0.

So, I must have a 0 0 is equal to 1 and within omega 0, it emits only one visible symbol which is b 0, so you find that the probability of emission from omega 0 that is b 0 of the symbol v 0 that will also be equal to 1. This is the only symbol which is emitted from this accepting state were as for all other states they have v 1, v 2, v 3. Similarly, from here I will have v 1, v 2, v 3 from here also I will have v 1, v 2, v 3 with the respective probabilities of emission.

So, here from omega 3 b 1 will be emitted with a probability b 3 1, similarly v 2 will be emitted with probability b 3 2 and so on. So, this is the final hidden Markov model that we are going to use for our temporal pattern recognition. Now, given such a hidden Markov model you find that in hidden Markov models, there are three central issues which need to be addressed, so what are the central issues.

(Refer Slide Time: 18:18)



In hidden Markov model are HMM the first issue or the first problem is an evaluation problem, the second problem is what is called a decoding problem. The third problem is which is very important not only in case of hidden Markov model, a very important problem in all the classifiers is the training of the classifier lining. Unless the training is done properly, the classifier cannot classify the patterns, so the third problem which is very important problem is the learning problem that is how the hidden Markov model is actually the patterns or how the hidden Markov model.

Now, what is this evaluation problem, now let us come to this problem first we said that when we have a hidden Markov model theta when this hidden Markov model theta is specified. We say that we have three things, one is the number of hidden states the number of whatever the hidden states we have the visible states or visible symbols. We have the transition probabilities a i j whenever the hidden Markov model makes a transition from state omega i to state omega j. We have the transition probabilities corresponding to the visible states though that actually probability, but all the visible symbols are also considered to visible states.

So, we can also call that as a transition probability, so it does not matter whether we call it transition probability or we call it a emission probability. So, we have to have the transition probability a i j for all a and for all j and we have to have the transition probability that is b j k or the emission probability that is b j k. This is the probability of emission of the visible symbol v k or visible state v k from the hidden state omega j. So, the model theta is specified by omega that is the set of hidden states v the set of visible states a i j for all values of i and j, which are the transition probabilities state transition probabilities and b j k which is the which are the visible symbol emission probability

So, once you have such a hidden Markov model theta and you have a sequence of visible symbols v t which is a sequence of length T. So, this evaluation problem is that given vt and theta we have to find out what is the probability that v t was generated by theta which is expressed. So, this is what the probability that this is visible sequence v t was generated by the hidden Markov model theta, but theta is specified over here. So, this we want to find out and this is represented as probability b t given theta and this is the problem which is evaluation problem.

Obviously once you evaluate this probability we can try to classify we can try to classify this visible sequence v t provided we know what this hidden Markov model is. So, you find that in case of hidden Markov model for recognition of temporal sequences if I have say have c number of sequences. An input sequence or an unknown sequence is to be classified to this to one of number of sequences, so I have c number of model sequences for every sequence I have to have a hidden Markov model theta. So, theta one the first hidden Markov model will encode or will represent the first sequence first model sequence model theta 2 will be for the second model sequence.

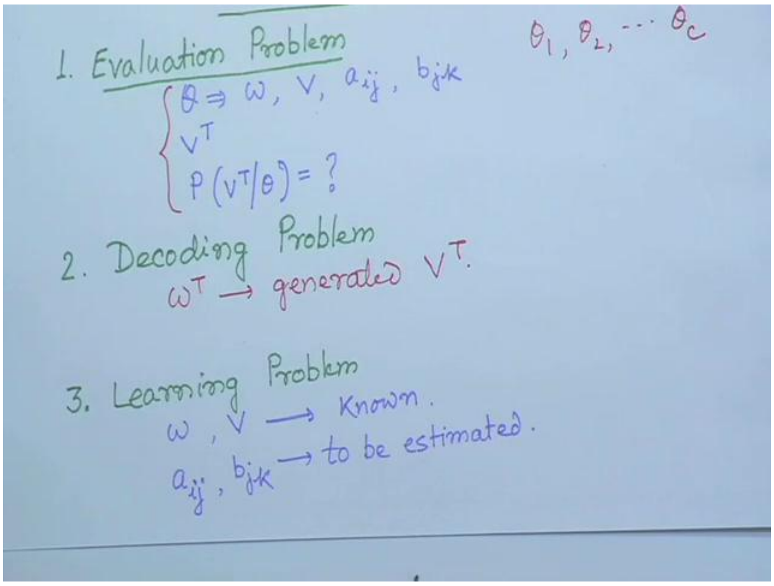
Now, theta three will be the third model sequence and so on, so once we have so many models from theta one to theta c theta 1 theta 2 or to theta c if I have c number of classes or if the input sequence is to be classified into one of the c classes. For every class, I have to have a Markov model, so every theta one we have corresponding that is the state of hidden states it will have a state of corresponding v that is the state of visible states. It will have corresponding a i j it will have corresponding b j k theta 2 will also have its corresponding omega a i j b j k and all that, so every sequence or every class is represented by a separate hidden Markov model.

So, given an unknown sequence v t we have to find out probability what is the probability that v t was generated by theta 1, we have to find out the probability what is the probability that v t was generated by theta 2. Similarly, what is the probability that v t was generated by theta c and from all this different probabilities then we can apply the base tool to classify the visible sequence b t. So, the evaluation problem actually deals with this that given a model theta and a v t a visible sequence symbols, we want to find out that what is the probability that v t was generated by this model theta, so this is what my evaluation problem.

Now, the decoding problems are that I have this v t and I have generated these probabilities what is the probabilities which was generated by the model theta. The decoding problems says that what is the most lightly sequence omega t of the hidden states which had led to generation of v t. So, we want to find out that sequence omega t that has laid to the generation of or which as generated this sequence v t that is what is the decoding problem and the learning problem is given a codes structure or half structure of the hidden model.

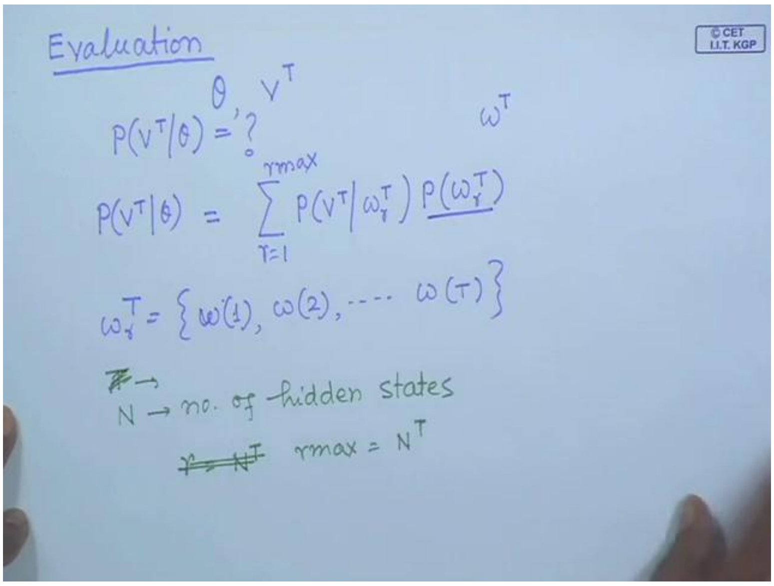
So, when I say the half structure means that how many hidden states this hidden Markov model has and how many visible states this hidden Markov model has. So, that is what is a half structure of a hidden Markov model, but I do not know the probabilities of transition the transition probabilities are unknown. So, through a set of given training sequences is I have to find out this transition probabilities, so I know what is omega that is the set of hidden states, I also know what is v that is the set of visible states.

(Refer Slide Time: 27:23)



So, these two are actually known what I have to find out is I want to know a set of training sequences a large number of training sequences is and using those training sequences, I have to estimate the transition probabilities a i j and b j k. So, these two transition probabilities are to be estimated. So, this is the problem which is known as learning problem or and as we said that this learning problem is one of them is a very important problem not only in case of hidden Markov model, but for designing any classifier. We will take up these central issues the evaluation problem, the decoding problem and the learning problem one by one.

(Refer Slide Time: 28:30)



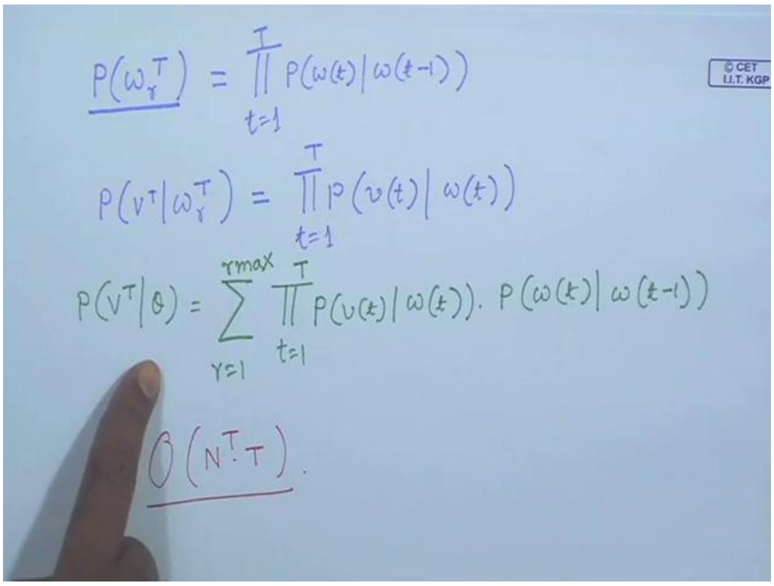
So, first let us take up the evaluation problem, so as we said that evaluation is nothing but given with a hidden Markov model theta and a sequence of visible states b t I have to estimate what is P v T given theta that is what the probability is. This model theta has generated this sequence of visible p t, now what I can do is I can find out all possible sequences of the hidden states of length T. So, that means I want to find out or I can that is the good approach all possible sequence that as we said that this v is nothing but sequence of t number of invisible states.

So, when I have a theta a number of visible states, I can find out all possible sequence of t number of visible states and for each of these sequence I can try to find out what is the probability that particular of hidden states omega T has generated this visible sequence. So, what I can say is in a good form this P v T given theta this can be estimated as I can find out what is P v T given say omega r t. Now, why writing omega r t am is i said that i want to find out all possible sequence of hidden states of length T this index r indicates one of those sequences it is one of those possible sequences.

So, I can find out what is P v T given omega r t that is one of the possible sequence which is to be multiplied by probability of omega t that is what is the probability. If I take the summation of this for r is equal to 1 to r max, were r max is the number of such a possible sequences that I can generate. So, if I take the summation r equal to r max of this, then what I get is P v T given theta that is the probability that this sequence of visible states v T is generated by this model theta. Now, over here this omega r t as I said that this is nothing but omega at times at one omega at time instant 2 continues omega at time instant T.

So, this is the sequence were for different sequences I have different values of omega 1 omega 2 and omega 3 and so on. Now, find that if there are say T number of hidden states if there are T number, let me put it because t values for something else. So, let me say if there are capital n number of hidden states this the number of hidden states, then you find that r max will be of the order of n to the power of t. So, this is r max is n to the power t, so I will have n to the power t number of possible sequences of hidden states of length T. So, this is what the value of r max and here I can compute you can compute that this p r of p omega r t because this is nothing but a sequence of states sequence of hidden states of length T.

(Refer Slide Time: 33:50)



So, I can easily estimate that p of omega r T this is nothing but p of omega T given omega t minus 1 from a state at time instant t from a state at time state t minus 1 it makes a transition to state omega T. This is the sequence of t number of such states this is actually a probability a transition of state at time state t minus 1 to another state at time state t. So, this is nothing but a probability of transition and from a particular sequence if I multiply this probability of transitions.

So, what I have to do is, I have to multiply this over t is equal to 1 to T, so if I take a product of this transition probabilities from t equal to 1 to T that gives me what is the probability of a particular sequence of hidden states. So, this p or t is nothing but this and these are nothing but transition probabilities, so this p r T p omega T is nothing but the product of the transition probabilities is from one state to another. At subsequent at consequent states for this r x sequence of invisible states or hidden states, similarly I can also compute p of v T given omega r T, you find that this is a visible symbol if I take a particular instant.

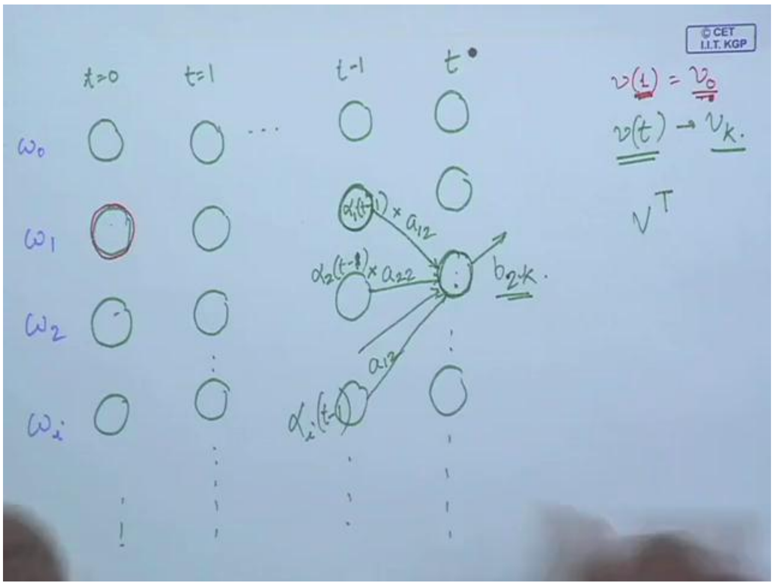
So, first instant the visible symbol the first visible symbol is emitted by the first state in the sequence omega r so this is nothing but the probability of emission of visible symbols. So, accordingly I can write this in the form this is the probability of emission of visible symbol like time step t given the machine is in emission at time step t. The product of this from t equal to 1 to T that gives me what is the probability b t given omega r t. You find that this is nothing but the product of b j k corresponding of b j k and every b j k will come from the hidden states omega T.

So, because of this two, I can now write this P v T given theta you find that P v T given theta was this expression of p v given theta. Now, in this expression if I put p of omega T or p omega v t given omega r t the expressions of which we have said that we have set the product forms. So, I can write p of v T given theta will be is equal to product P v T given omega t into p omega t given t minus 1 say just look at this expression and this expression. So, this is nothing but this, so I have to do this take this product is equal to T and to take the summation of this over r is equal to 1 2 r max.

So, given a hidden Markov model theta and a sequence of visible states v t, I can find out the probability that this model theta has generated with this sequence of visible symbol because v t as this expression. Now, if I look at the complexity of this particular expression the complexity of this expression will be at the order of n to the power t time t which is huge. So, I have to have a high number of computations a very large number of computations to find out this P v T given theta.

So, instead of using this direct expression for computation given theta, I can make use of a simpler approach. I can make use of algorithm the algorithm which will use that given a sequence of visible symbols v t what will be the probability that the hidden Markov model will be in a particular state at a particular state and time. So, how is that possible let us just try to explain it conceptually, so I have a number of states is as I said one of the states is that I have is an accepting state or an absorbing state or a final state omega 0.

(Refer Slide Time: 40:09)



So, I have state omega zero omega one omega two like this I have omega i continues has a number of states, so I will put it like this these represents states and I have like different time states I will have same number of states. So, these is a time state 0 is equal to 0, this is at t equal to 1, similarly I will have this is t minus 1 I will have time state t it continues like this. So, the first row indicates omega 0 second row is omega 1 omega 2 say omega i some are like omega j. So, you find that if somehow i know that at equal to zero what is the hidden state in which the machine.

So, let me assume that the machine is in hidden state omega 1, then I can find out that at t equal to 1 what is the probability that the machine state in in state omega 0 what will be the probability that the machines will be in state omega 1. What is the probability that the machine will be in state omega 2 what is the probability in the machine state omega i and so on. The reason is from omega 1 it can make a transition possible to make a transition to omega 0, it is possible to make a transition to omega 1 omega 2 omega i like this.

So, if v t in v T the symbol v 1 if the symbol v 1, so b one the symbol v 1 is equal to v 0 which is the visible state emitted by the final state only. So, if this is equal to this, then obviously from omega 1 I will have a transition to omega 0 because it is this only omega 0 which can emit this v 0 and it times step 1, I have this visible symbol 0. If this is the case, then from omega 1, I definitely have a transition to omega 0 if this is not the case if v 1 is not equal to v 0. Then obviously from omega 1, the model has not made a transition to omega 0 at time state t 1 it has a transitions to some other states, now let us consider that any of the state.

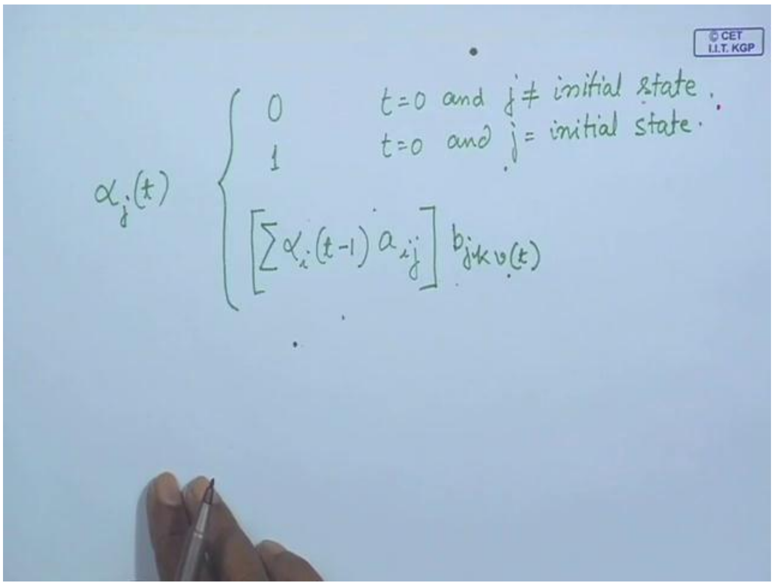
So, I have said this time state t and so on, so let us consider this particular state omega two so what are possible ways that in which the machine can come to this state omega 2. Obviously, it cannot make a transition to omega 1 to omega 2 because as we said that this model enters the final state omega 0, it cannot come out of that state. So, from omega 0, it cannot make a transition to this this state omega 2, however it can make a transition to omega 2 to omega 1; it can make a transition to omega 2 to omega 2. It can make a transition from omega 3 to omega 2, it can make a transition omega i to omega 2 and so on.

Now, if somehow I know what is the probability I have been able to compute the probability that I have been machine state this omega 1 at time instant t minus 1 and that is given by alpha 1 t minus 1. Here, it is alpha 2 t minus 2, here it is alpha i t minus 1, so this the probability that the machine is in these states at time instant t minus 1. Then I also know that what is the probability at the machine will take a transition from omega 1 to omega 2 which is given by the transition probability a 1 2. So, the probability that the machine can make a transition from omega 1 to omega 2 in time step t is given by this alpha 1 t minus 1 into a 1 2.

I also have a possibility that machine will make a transition to omega 2 to omega 2 itself and this probability transition say 2 to 2. So, the probability that the machine makes a transition from omega 2 to omega 2 is nothing but alpha 2 minus 2 into t minus 1 into omega 2 into a 2 2. Here, it is alpha 1 t minus 1 into a 1 2, here it is alpha i t minus 1 into a i 2 and here if I make the sum of all these products what I get is the probability that there will be a transition from one of the states. From any of the states in time state t minus 1 to state omega 2 in time state 2 and I also know what is B t, that is the visible symbol in time state t.

This visible symbol time step t can be emitted by this state omega 2 with a probability b 2 and suppose this emitted symbol is v k. So, this will have a probability b 2 k, so finally I can say that the probability that the machine is in state omega 2 in time state t. After emitting the first t time visible symbols from i visible sequence of states v t is given by some of all this product terms multiplied by this b 2 k. So, by using this logic I can simplify or I can have recursive algorithm to find out the probability that the sequence v T was generated by hidden Markov model theta.

(Refer Slide Time: 47:42)



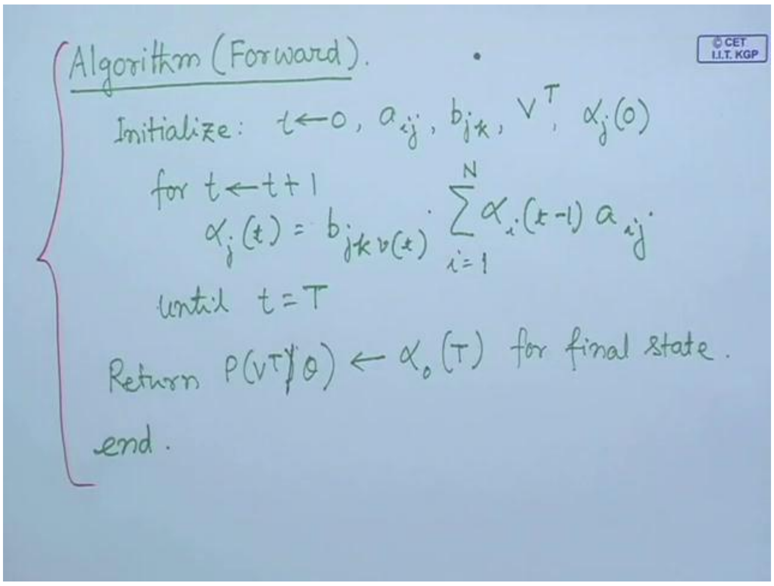
So, for that what I do is I define the term say alpha j t which as we have just explained that this alpha j t this simply says what is the probability that the machine will state omega j in times step t. After emitting first t number of visible symbols in the sequence of visible symbols v of T, so that is what and we can define that this alpha j t will be equal to 0. If t equal to 0 and j is not an initial state t equal to 0 means we are talking about the initial condition. As per this diagram and the probability as I said that if I know that the machine is initially in state omega 1.

So, I am saying that this is equal to the probability of omega one time state t is equal to 0 1 is 1, the probability that the machine will be in any other states will be equal to 0. So, the as per definition I said omega j t is equal to 0 if I am in the initial state that is t equal to 0 and this j is not an initial state. If it is an initial state then this will be equal to 1 that is t equal to 0 and j is initial state, otherwise I define this alpha j t as some of alpha i t minus 1 into a i j into b j k b t. You come to this diagram that we have drawn this is this is the product the sum of all this products this is equivalent to sum of alpha t alpha i t minus 1 into b i j that is the transition probability from the i th state to the j th state times b k j v t.

So, what we have said here earlier is v t is equal to v k then I take the initiation probability to transition probability to b 2 k that is from the j th state at time t whatever symbol is emitted. I choose only that corresponding probability to decide what is the probability that can be in state omega j at time step t because I know what is the t th visible symbol in my sequence of visible symbols of this, I can make use of that.

So, this is what I have, so this b j k v t indicates that this b j k is the probability of transition is chosen by the t th symbol, so if b t is equal to v 1 this will be simply b j 1 if v t is equal to v 4 this will be b j 4. So, I make use of this particular definition and using this definition, now write I can write an algorithm to find out what is the probability that the machine theta the HMM theta has generated this sequence of visible symbols v T.

(Refer Slide Time: 52:03)



So, I write an algorithm and in particular I call it as forward algorithm the reason is why am writing forward is ill make use of a similar algorithm that is call a backward algorithm. Both this forward algorithm and backward algorithm will be used in a third issue or learning of the hidden Markov model. So, there will use forward and backward algorithm both of them together I will simply write this forward algorithm, so obviously the first one is initialization state.

So, I have to make time state t is equal to 0, I have to know what is a i j, I have to know what is b j k, I have to know what is v T because actually this is the one I am trying to find out the probability and I have to initialize alpha j 0. So, this depends upon if any initially assume that the machine is in state omega 1 alpha 1 0 will be equal to 1 alpha 2 0 will be equal to 0 alpha 3 0 will be equal to 0 and so on. Then I go for what I do is for t increment t alpha j t will be b j k v t the same one into some of alpha i t minus 1 into a i j were i varies from 1 to n as I have n number of hidden states.

This has to continue until t becomes to T because I have to take all the symbols from given sequence of symbols. So, this has to continue for till t becomes to T. At the end of this what I have to do is I have to return P v T, P v T given theta which is nothing but alpha 0 t and this alpha 0 is the probability of final state. So, this is for final state and the algorithm stops here.

So, this is the algorithm which can be used called forward algorithm used to find what is the probability that a given sequence v T is generated by a given model theta and while doing so we have also found out that in every step what is the probability that the machine can be given in a particular hidden state. So, will stop this lecture now, in the next lecture ill explain this algorithm further with an example.

Thank you.