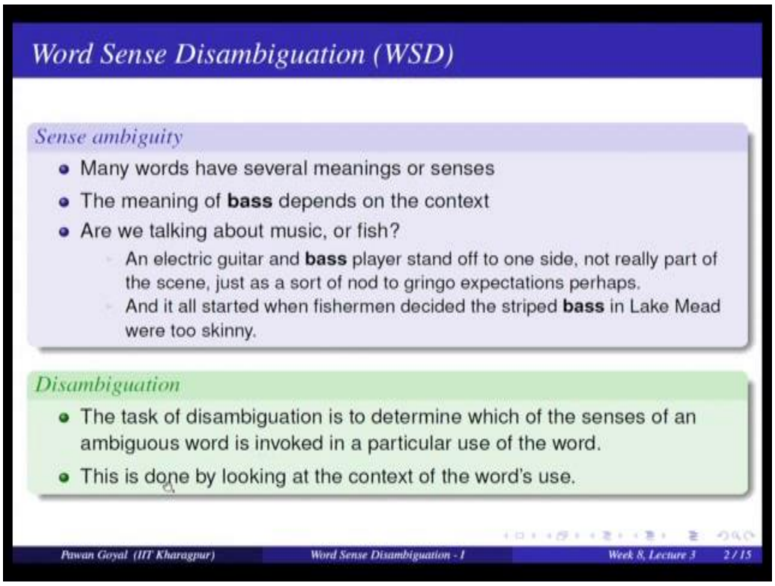
Natural Language Processing
Prof. Pawan Goyal
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 38
Word Sense Disambiguation – I

Welcome to the third lecture of this week. So we are talking about lexis semantics and there we dealt with what are the various relations we can establish between the word forms. So we were talking specific in traditional now that are lexemes different items in my lexicon. And we also talked about word net. That what are the different relations we can establish between various lexemes word net. And what are the different standard forms like using sins at and what is the hierarchical award net, how do you use the define word similarity and so on. And finally, we ended up the saying that there is one in head problem engaging word net that is when you get a new sentence you do not know what is the particular sense of the word that is being use in the sentence. So you cannot apply the methods vary directly.

So that is fair. So we deal with this problem that given a sentence for each word find out what is the sense in the word net that is being used here. And this problem is called versions disambiguation. So very classical problem in NLP and let lot of research has happened in over last few decades, so we will talk about what are the most important methods for dealing with this problem. And we will talk about both simple approaches and machine learning based approaches and also some unsupervised approaches. So we will cover that in the next 2 lectures starting today. So this is our topic that is word sense disambiguation.

(Refer Slide Time: 01:55)



Word Sense Disambiguation (WSD)

Sense ambiguity

- Many words have several meanings or senses
- The meaning of **bass** depends on the context
- Are we talking about music, or fish?
  - An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.
  - And it all started when fishermen decided the striped **bass** in Lake Mead were too skinny.

Disambiguation

- The task of disambiguation is to determine which of the senses of an ambiguous word is invoked in a particular use of the word.
- This is done by looking at the context of the word's use.

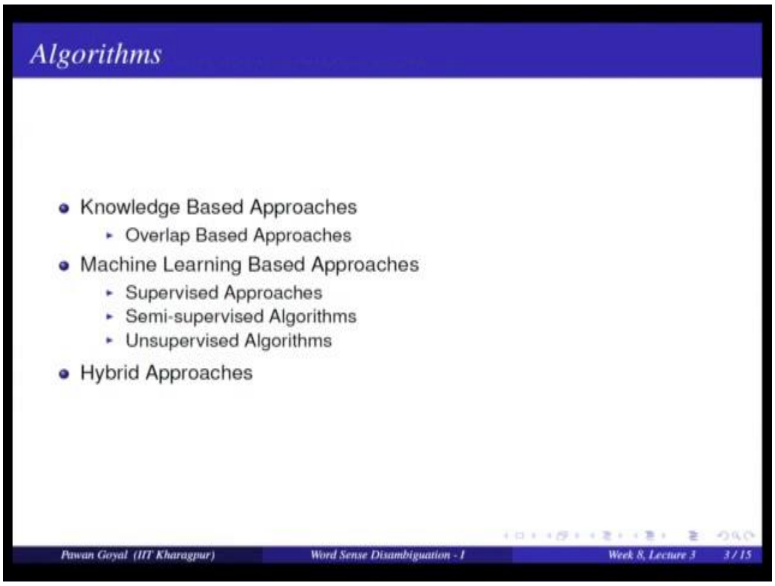
Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lecture 3 2 / 15

So let me just define the problem again. So we have seen in this week that there are many words that are having several meanings. And then we can call them different senses of the same word. So now, example can be of the word bass. So you can use bass or base depending on that they you want to use that in the context of fish or music. Now the problem is whenever you are given this word I will be talking about music or fish, you know. So let us see these 2 sentences here. The first sentence is an electric guitar and best bass players stand off to one side not really part of the scene just as a sort of nod to gringo expectation perhaps. This is one sentence. And the second sentence is talking about some trip bass in Lake Mead were too skinny. So now, the same word is being used in both the sentences. And I have to find out whether the word is being used the sense of fish or music. Now once you see the sentences, how do you find out that what is the meaning of this word in the sentence. You will try to look around the context in what context this word is being talked about. Accordingly, you will choose one sense over and another one. Now this whole the field of words has disambiguation.

So this deal with this problem, that I have to find out the sense of the word depending the context know. How can I do that computational? So we can define the task of the disambiguation that is to determine which of the senses of an ambiguous word is invoked in a particular use of a word. So we say we are doing this problem only for the ambiguous words. Whenever it has more than one senses in a particular use of that word, what is the sense that has been used.

Now, and as you can see we will deal with this problem by looking at the context of the words use.

(Refer Slide Time: 04:01)



Algorithms

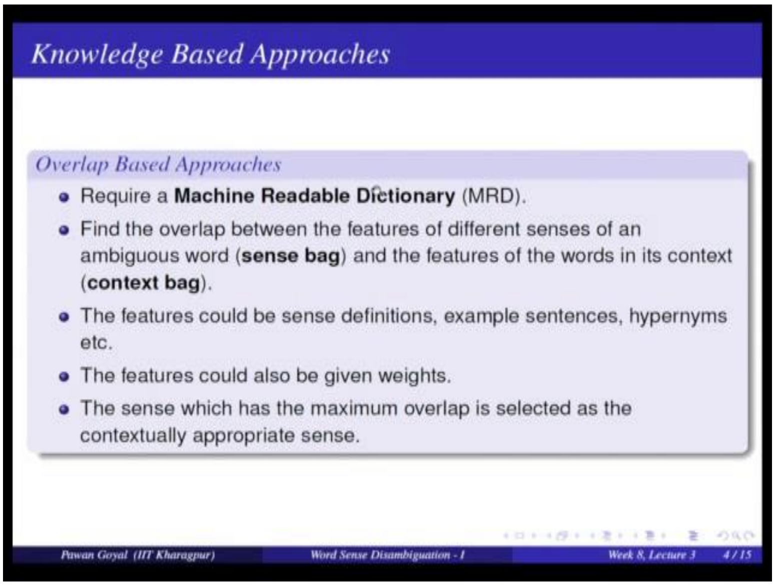
- Knowledge Based Approaches
  - Overlap Based Approaches
- Machine Learning Based Approaches
  - Supervised Approaches
  - Semi-supervised Algorithms
  - Unsupervised Algorithms
- Hybrid Approaches

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lecture 3 3 / 15

So now how exactly we do that? There are many different algorithms that handle this problem. So there are approaches based on their knowledge base approaches that use different overlap based methods, for handling that. Then there is some machine learning based approaches they use supervised approaches unsupervised approaches and semi supervised approaches. And also there are some hybrid approaches. And in the literature there are a lot of different methods for solving this problem.

Now, what we will do? We will talk about some very basic methods. So that you have the intuition with you that what are the different features what are different methods you can apply for word sense disambiguation.

(Refer Slide Time: 04:39)



Knowledge Based Approaches

Overlap Based Approaches

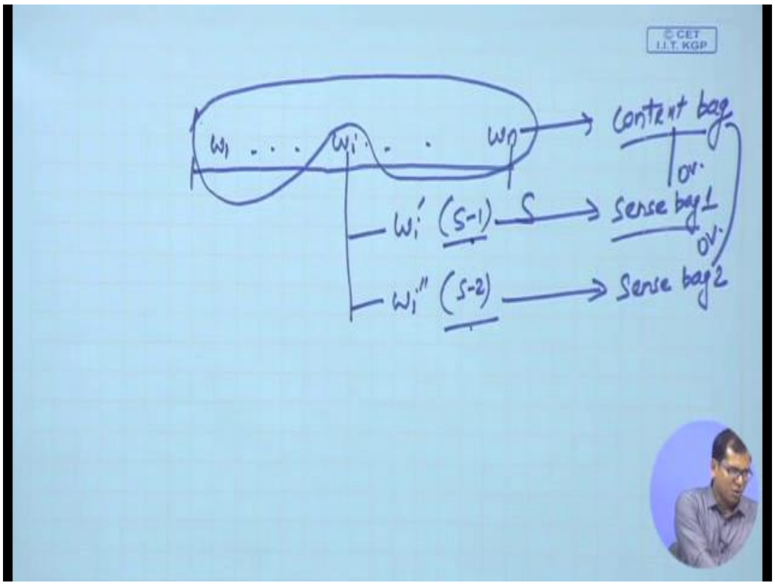
- Require a **Machine Readable Dictionary** (MRD).
- Find the overlap between the features of different senses of an ambiguous word (**sense bag**) and the features of the words in its context (**context bag**).
- The features could be sense definitions, example sentences, hypernyms etc.
- The features could also be given weights.
- The sense which has the maximum overlap is selected as the contextually appropriate sense.

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lecture 3 4 / 15

So, now starting with the knowledge base approaches, so there are some sorts of overlap based approaches. They try to use some sort of overlap. Now how do they find is overlap? They would require a machine readable dictionary. So like word net is my machine readable dictionary. You can have some other sort of the torus also. That are in a machine readable form. So by machine readable I mean all the entries and all the information is such that you can easily access that and you can make use of that.

So what approaches will do? They will try to find the overlap between the features of different senses of an ambiguous word that is they will call it sense bag and the features of the word in it is context, and call it as the context bag. And they will try to measure the overlap. So now, just to explain this idea let us see I am having a sentence.

(Refer Slide Time: 05:34)

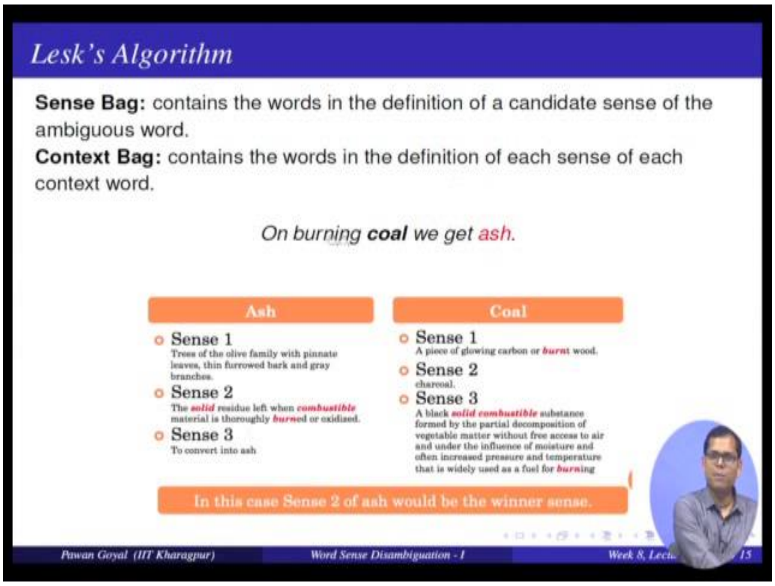


S, that contains some words w1 to wn. And the word w i can be used in 2 senses. W i prime and w i double prime; it is sense 1, sense 2. Now in the sentence this word w i would be used only for one sense and that is what we are assuming. This will be true in most of the cases unless, the speaker himself wants to imply 2 different meanings of the same words. So we will say in general for this word w i you would have only one sense.

So now I want to find out depending onthis context whether sense one should be taken off sense to should be taken. So idea would be for each sense I will construct the sense bag. And we will see how do we construct sense bag. So this I will convert that to a sense bag. Sense bag 1 and this I will convert to sense bag 2. Now looking at the context that is you can think of all the words that are coming in the sentence other than this word I will construct a context bag. Now I will try to find out what is the overlap between context bag in sense bag 1 and what is the overlap here. And I will take the answer as the as the one sense that is having the highest overlap. And this is to explain that simply, I have various senses, let us context the sense bag context that the context bag take the overlap take the sense with the maximum overlap.

Now, to construct this sense bags and context bag, you will use various features that are provided in the dictionary that you have. So features could be like what are the sense definitions that I am having, what are the example sentences that are provided, different hyponyms and so on. And any other criteria that you can use from the dictionary and you will use that to construct both the sense bags and the context bag take the one with does maximum overlap.

(Refer Slide Time: 07:53)



Lesk's Algorithm

Sense Bag: contains the words in the definition of a candidate sense of the ambiguous word.
Context Bag: contains the words in the definition of each sense of each context word.

*On burning **coal** we get ash.*

Ash

- Sense 1
  Trees of the olive family with pinnate leaves, thin furrowed bark and gray branches.
- Sense 2
  The ***solid*** residue left when ***combustible*** material is thoroughly ***burned*** or oxidized.
- Sense 3
  To convert into ash

Coal

- Sense 1
  A piece of glowing carbon or ***burnt*** wood.
- Sense 2
  charcoal.
- Sense 3
  A black ***solid combustible*** substance formed by the partial decomposition of vegetable matter without free access to air and under the influence of moisture and often increased pressure and temperature that is widely used as a fuel for ***burning***

In this case Sense 2 of ash would be the winner sense.

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lect 15

Now, let us take a simple example or in actual algorithm that that does that. So we have lesks algorithm. So we talked about this algorithm in the previous lecture also. For a different case when you wanted to find out what is some ready between 2 different senses. Now suppose I want to use this this less algorithm for the purpose of versus disambiguation. So what it does? So it constructs the sense bag context bag and how are they constructed? The sense bag contains the words in the definition of the candidate sense of the candidate part.

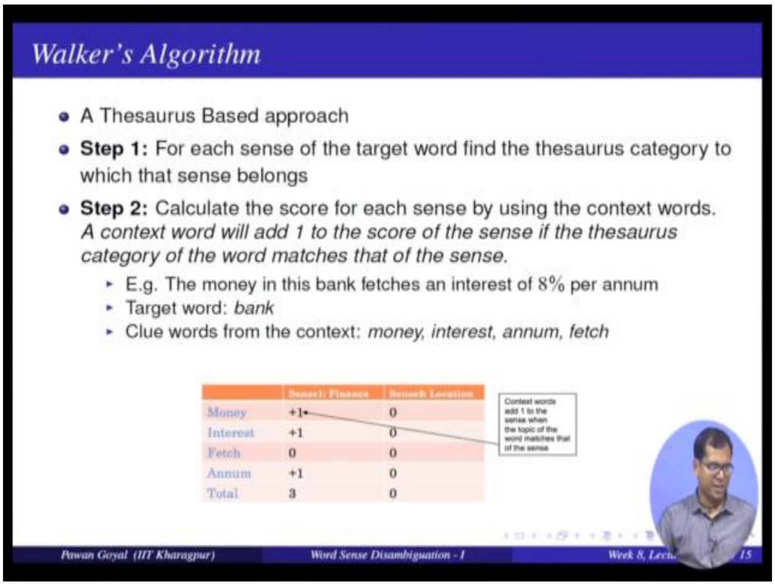
So I have a word that it ambiguous can have multiple senses, for each of the senses which I will construct one sense bag. And the sense bag is constructed by using the definitions that I have in my dictionary. And how do I construct the context bag? I take all the words in my context, take each of the senses of the context word and then take their definition all these together become my context bag. So I have a single context bag and I have multiple sense bag for different senses of the word. And then I take the overlap and this overlap is taken by using the leska algorithm similar to what we saw in the previous lecture, that if there is a match of n a particular n gram I a d score of n square.

So now let us take this case. So I have a sentence on burning coals we get ash. Now hear the word ash it ambiguous and I want to find out what is the correct sense of this word that is usually sentence. And suppose for simplicity we are using only the word coal as my context. So what am I going to do? I will find out, I will make different sense bag for the word ash whatever senses are recorded in my word net, and I will construct the context bag by using all the definitions of the word coal. And then I will try to measure the overlap. So suppose we do that, so here is the case from by taking the dictionary definitions from word net. So ash has 3 senses. So one is corresponds to some trees second corresponds to some solid residue and third is to converting to convert that into ash. And we want to find out what which of the 3 senses is used here.

So what we will do? We will take the word coal that is the only context for we are using here, but in general you can use any number of context words. I take all it is sense definitions that are defined combine them together to make a context bag. So this whole thing together becomes my contacts bag. Now I measure it similarity with each of the senses. And if you do that we find the sense to 3 of the words are matching here and sense to become a winner.

Here, what you using here? You are also using that is stemming. So this is optional you might use stemming you may not use stemming. So this is the generic idea of leska algorithm. I hope this is clear, if I want to use multiple words I can repeat the same thing. I suppose I am using burning coal and get all the 3 words, I will take all their definitions and combine them together into a single context bag. And then measure the similarity of each of senses with this whole context bag.

(Refer Slide Time: 11:11)



Walker's Algorithm

- A Thesaurus Based approach
- **Step 1:** For each sense of the target word find the thesaurus category to which that sense belongs
- **Step 2:** Calculate the score for each sense by using the context words. *A context word will add 1 to the score of the sense if the thesaurus category of the word matches that of the sense.*
  - E.g. The money in this bank fetches an interest of 8% per annum
  - Target word: *bank*
  - Clue words from the context: *money, interest, annum, fetch*

	Sense1: Finance	Sense2: Location
Money	+1	0
Interest	+1	0
Fetch	0	0
Annum	+1	0
Total	3	0

Context words add 1 to the sense when the topic of the word matches that of the sense

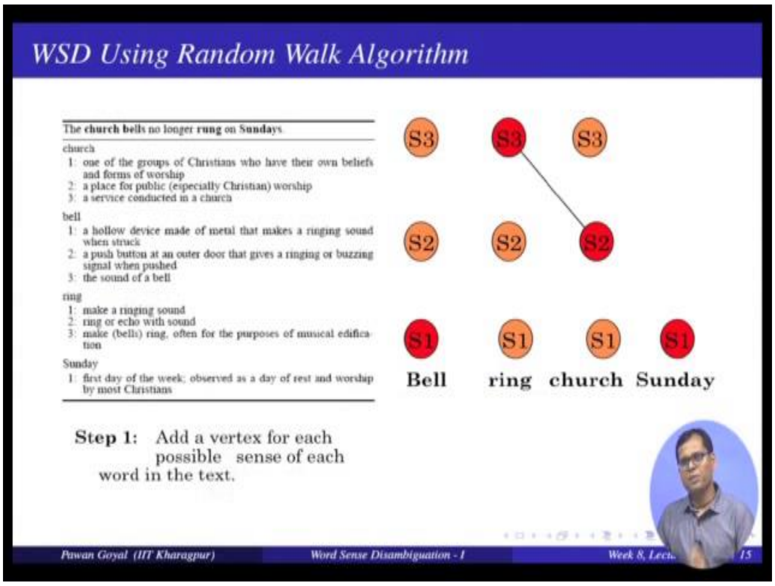
Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lect... 15

So now let us take another scenario, where we have a different kind of source. So what is this source now? For each word we are told what are the categories these senses it is senses belong to. So I have some finite set of categories. This can be like finance category, location categories, sports category and so on. And each word has been recorded as it has a sense in these this category. So now, the problem would be when I am given a sentence a word can have senses in multiple categories find out what is the a probably category in this sentence. So how do we use this framework here? So what I will do, for each sense of the target word if find out what is the source category to which this belongs; now calculate the score for each senses by using context words. No idea would be the context words again would be having the researches all categories. So for each context, but I will find out what is the source category over I will see by using the context word which thesaurus category getting the highest count.

So here, what we will be doing? A context word will add score of 1 of the sense if the thesaurus category of the word matches that of the sense. So let us take this sentence the money in this bank fetches interest of 8 percent per annum, and suppose my ambiguous word here is bank has 2 senses. One in the case of safe finance that is for the money bank that is also used in the sentence another could be river bank also can be used as location or something. And I take each context words, money interest and annum fetch. Each again would have their own dissolves category.

So what we will be doing? I will take all these words money interest fetch annum. I know bank have 2 different senses finance and location. So for each of these context words I will put a one if the match any of this, this all I can take as much. So money message finance, so I will put a one here, interest message finance again I will put a one here, fetch does not meet any of these senses. So both are 0 and a message finance again plus 1. And then I will add the score of both the senses and I in here I say that a finance the score is 3 for locations score is 0. So I will take the sense of bank is finance in this particular case. So these are simple approach it by just taking the dictionary definition thesaurus category.

(Refer Slide Time: 13:49)



WSD Using Random Walk Algorithm

The **church bells** no longer **rung** on **Sundays**

church
1: one of the groups of Christians who have their own beliefs and forms of worship
2: a place for public (especially Christian) worship
3: a service conducted in a church

bell
1: a hollow device made of metal that makes a ringing sound when struck
2: a push button at an outer door that gives a ringing or buzzing signal when pushed
3: the sound of a bell

ring
1: make a ringing sound
2: ring or echo with sound
3: make (bells) ring, often for the purposes of musical edification

Sunday
1: first day of the week; observed as a day of rest and worship by most Christians

Step 1: Add a vertex for each possible sense of each word in the text.

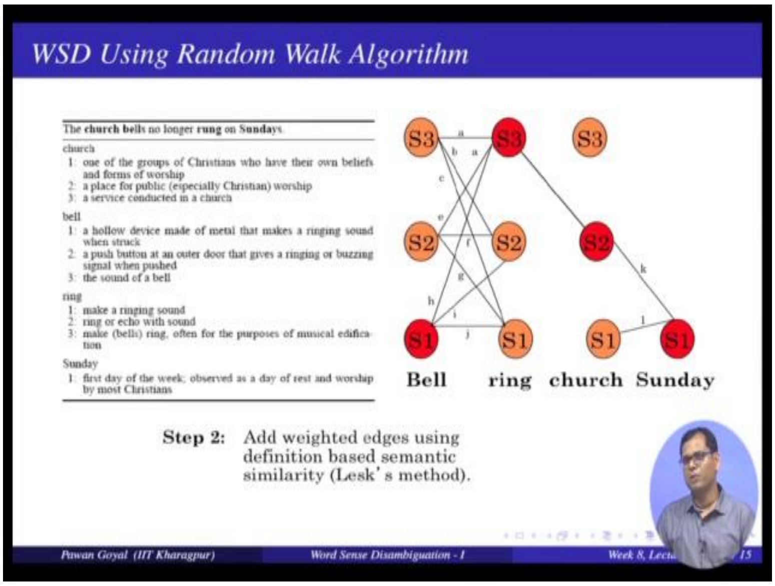
Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lectu 15

So there is one problem in both these approaches. That is work for disambiguating the sense of a particular word we are not disambiguating the sense of the other context words. We are taking all the possible senses of them into a single context bag, but suppose you want to solve this problem where in the same sentence there are multiple ambiguous words and you want to disambiguate each of them together. Then can we do that? And this can be done by making each of the fact that although each word might have multiple senses, but the particular senses that are being used in this sentence there will be somehow connected to each other.

So we want to exploit this and for this we can use some sort of grab based method. Similar to what we do in the case of PageRank algorithm. So let me just quickly explain this idea. And we will actually deal with this PageRank algorithm in detail when we talk about summation application, but I will try to give you the intuition. So this is the problem that suppose I have the sentence, the church bells no longer ring on Sundays. And I can say there are 4 words bell rings church in Sunday. And I want to disambiguate the senses of these words.

And also suppose that from the dictionary I know what are there sense definitions. So here the word church has 3 senses, bell has 3 senses, ring has 3 senses, and Sunday has one senses. And I want to disambiguate the first 3 words together. So how do I do that? So idea would be that is treated as a graph based problem, where for each word I first write down what are different senses of this word. So I have made a vortex for each of the sense of these words. So there are 3 words is for a bell 3, for ring 3, for church and one for Sunday. So I have now 10 words in this graph. Now what will be the idea? I will now try to connect the different words together by seeing what is the overlap among different sense definition.

(Refer Slide Time: 16:05)



WSD Using Random Walk Algorithm

The **church bells** no longer **rung** on **Sundays**.

church
1: one of the groups of Christians who have their own beliefs and forms of worship
2: a place for public (especially Christian) worship
3: a service conducted in a church

bell
1: a hollow device made of metal that makes a ringing sound when struck
2: a push button at an outer door that gives a ringing or buzzing signal when pushed
3: the sound of a bell

ring
1: make a ringing sound
2: ring or echo with sound
3: make (bells) ring, often for the purposes of musical edification

Sunday
1: first day of the week; observed as a day of rest and worship by most Christians

Step 2: Add weighted edges using definition based semantic similarity (Lesk's method).

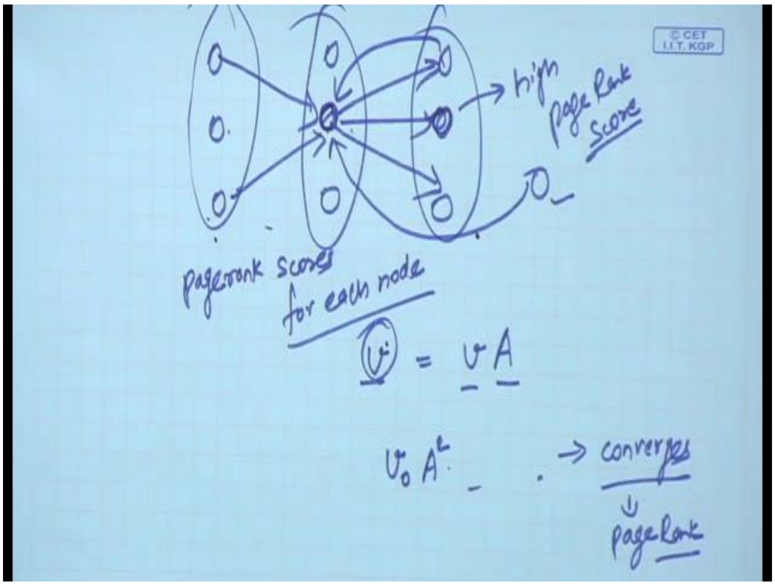
Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lecture 15

So I will now try to connect some weighted edges between different sense definitions by using Lesk's method. So the Lesk's method is by similar by simple for finding similarity between 2 sentences. I use that find out what is the similarity between any 2 sense definitions, and add a weight between these 2. And this will give me some sort of matrix like that. So I have all the sense definitions connected. Now one thing one important thing is that for the same word I will not have to connect it is own sense definition. So there will not be any edge between S3 and S2 for the same word bell, but they will be edges between different words. So S3 of bell an S2 of ring would be connected by what is the similarity between these 2 sense definitions.

So idea is that now we want to somehow find out that particular combination here. So one sense for each of the 3 words such that their overall similarity is the highest. So they are having the highest similarity. And one particular method that can use for that is using it a PageRank kind of algorithm. So idea would be, if there are multiple, if there are appropriate sense definitions that are similar to each other, if I use the PageRank algorithm, they would have very high ranking school because they will all contribute to each other.

So what do we do? So for this graph once we have all the word decease we have all the edge bit also, we treated as a problematic fear finding the PageRank for each vertex of this graph and to give you simple intuition.

(Refer Slide Time: 17:56)

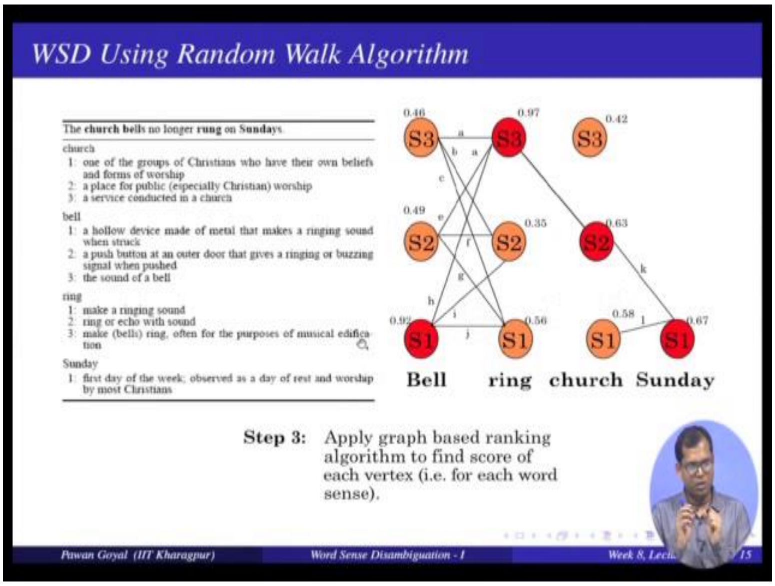


So now, what would we have? So we have 10 nodes the graph and I try to connect them by different numbers. These numbers depend on their lesks similarity. Now once I have this if I want to compute the PageRank scores for each node. And now what is the algorithm for computing paging is score if you take it the simple idea that is I suppose this PageRank it completed in a creative manner. So we start with some initial random is course, this will be some sort of probable distribution, and I use this equation v equal to v A to find out what will be the PageRank score and we denotes the PageRank scores. And how do we actually come up with this score? Start with some initial v 0 and then keep on multiplying a square a cube until it converge and that becomes your PageRank score.

So we will talk about this algorithm in detail later and for now you can just quickly have a look at what is the PageRank algorithm, but the idea would be once we do all that you will find one PageRank score for each node. And the nodes that are having a lot of connections with other nodes and with high in degree connection will be given a higher PageRank score. So if suppose this node is connected to multiple sense definitions. It is connected to this sense definition, this sense definition, and this sense definition; that means, this might be one of the important senses in this for this particular word. And idea would be to find out for each word one sense that is having the highest connection and this would be someone that will get a high PageRank a score.

So I will give PageRank the score to each of the 10 node and then for each word I will pick the one that is having the highest among this set. So I will take one from each of the set and here I anyway choose this one. And this becomes my final disambiguated sense. So once I have added all these weighted adages, I will apply the graph based ranking algorithm, so this can be PageRank algorithm.

(Refer Slide Time: 21:00)



WSD Using Random Walk Algorithm

The **church bells** no longer **rung** on **Sundays**

church

1: one of the groups of Christians who have their own beliefs and forms of worship
2: a place for public (especially Christian) worship
3: a service conducted in a church

bell

1: a hollow device made of metal that makes a ringing sound when struck
2: a push button at an outer door that gives a ringing or buzzing signal when pushed
3: the sound of a bell

ring

1: make a ringing sound
2: ring or echo with sound
3: make (bells) ring, often for the purposes of musical edification

Sunday

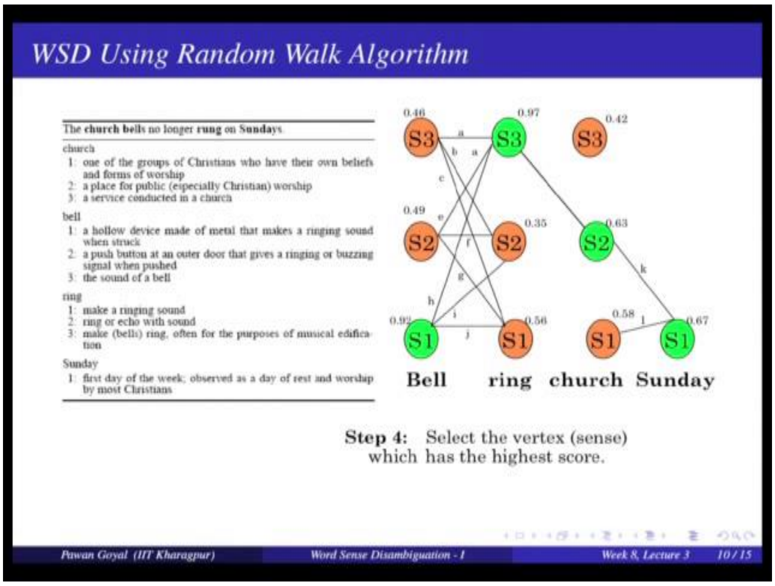
1: first day of the week; observed as a day of rest and worship by most Christians

Step 3: Apply graph based ranking algorithm to find score of each vertex (i.e. for each word sense).

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lec... 15

Now once we get all this course I will for each word I will choose the one with the highest score. An idea here is the sign the purpose sense that is connected to multiple other sense which for multiple words would get a higher score. And that is the intuition that I want to choose the particular sense that is having a similar definition has many other senses. And this I do for each of the word. So here suppose S1 in for bell, one S3 for ring and S2 for church are selected and they are my final disambiguated senses for these 3 words. And this is my overall algorithm.

(Refer Slide Time: 21:47)



WSD Using Random Walk Algorithm

The **church bells** no longer **rung** on **Sundays**

church
1: one of the groups of Christians who have their own beliefs and forms of worship
2: a place for public (especially Christian) worship
3: a service conducted in a church

bell
1: a hollow device made of metal that makes a ringing sound when struck
2: a push button at an outer door that gives a ringing or buzzing signal when pushed
3: the sound of a bell

ring
1: make a ringing sound
2: ring or echo with sound
3: make (bells) ring, often for the purposes of musical edification

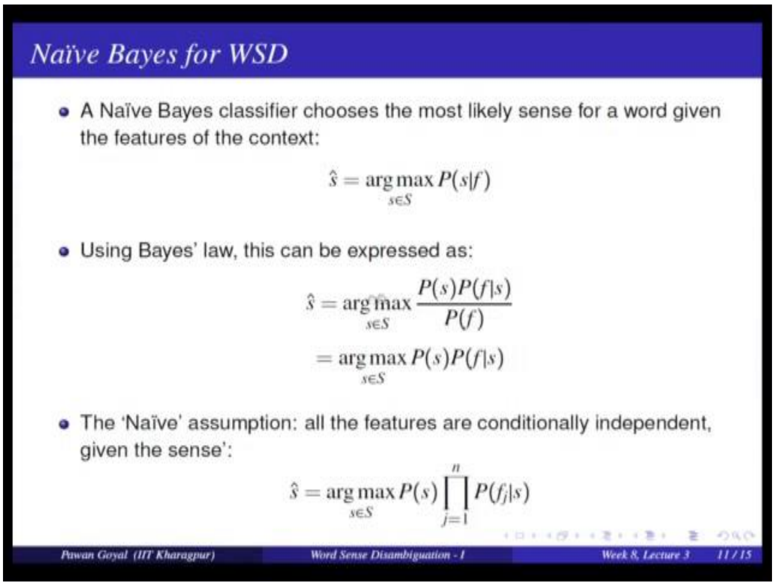
Sunday
1: first day of the week; observed as a day of rest and worship by most Christians

Step 4: Select the vertex (sense) which has the highest score.

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lecture 3 10 / 15

Now, we can also use some machine learning based methods. And one simple idea would be to use a Naive Bayes algorithm.

(Refer Slide Time: 21:53)



Naïve Bayes for WSD

- A Naïve Bayes classifier chooses the most likely sense for a word given the features of the context:

$$\hat{s} = \arg\max_{s \in S} P(s|f)$$

- Using Bayes' law, this can be expressed as:

$$\hat{s} = \arg\max_{s \in S} \frac{P(s)P(f|s)}{P(f)}$$
$$= \arg\max_{s \in S} P(s)P(f|s)$$

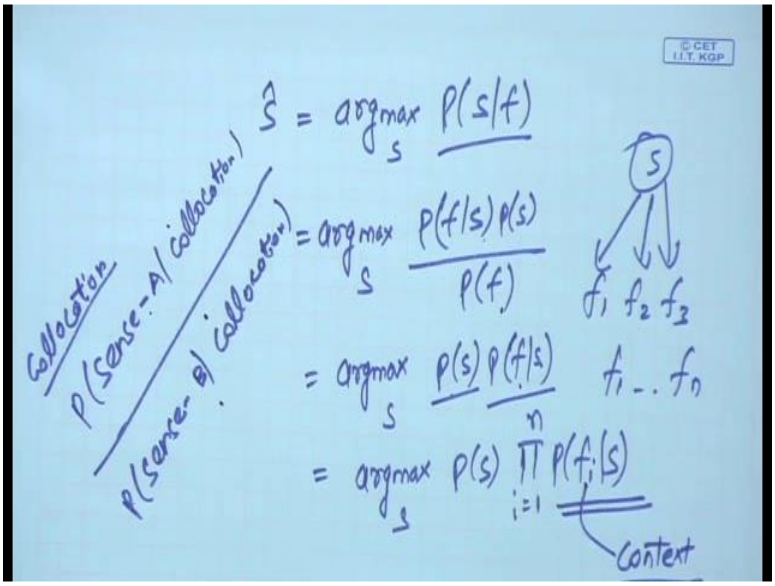
- The 'Naïve' assumption: all the features are conditionally independent, given the sense':

$$\hat{s} = \arg\max_{s \in S} P(s) \prod_{j=1}^{n} P(f_j|s)$$

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lecture 3 11 / 15

So Naive Bayes algorithm we are using for doing classification here, what is the classification task for a given word there are multiple senses, from the sentence find out what is the appropriate sense. So this method has to be word specifics and for each word you might have as many classes as the number of sense of this word.

(Refer Slide Time: 22:37)



Collocation

P(sense-A | collocation)

P(sense-B | collocation)

$$\hat{s} = \arg\max_s P(s|f)$$

$$= \arg\max_s \frac{P(f|s)P(s)}{P(f)}$$

$$= \arg\max_s P(s)P(f|s)$$

$$= \arg\max_s P(s)\prod_{i=1}^{n} P(f_i|s)$$

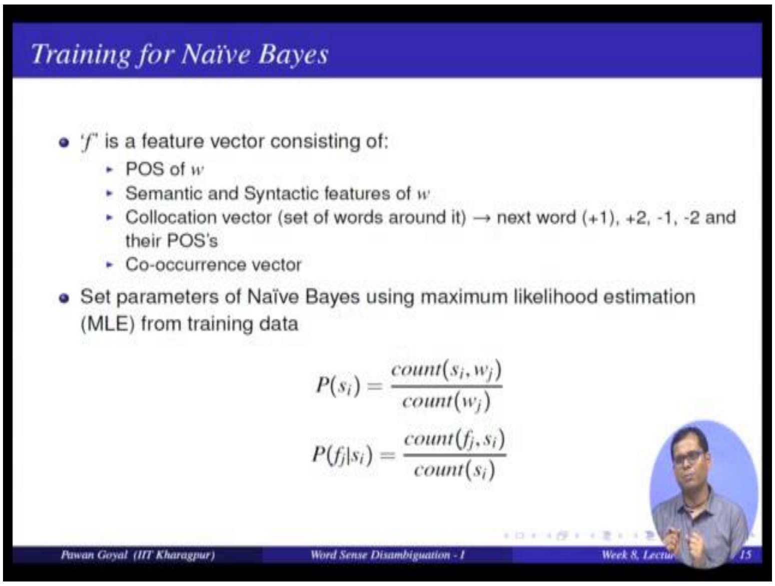
So problem here is, find out the sense of the word that gives the maximum probability as given f and we have a set of features that I will extract from the context around this word. So this for a given there can be multiple senses I want to find out the sense that is having this score at maximum. Now because Naive Bayes genetic model; that means, the sense comes on the off first and then the features are generated from their different features. So how do we compute this particular probability? This would be nothing, but orgmax s probability f given s probability s divided probabilities f and because this common for all the sense this is orgmax over s P s, P f given s is now P s nothing, but the prior probability of the sense. That is if a particular sense of the word is more commonly used that many will get a high prior and P f given s h what is the probability of different features being observed or generated by this sense.

So Naive Bayes model what we do? We make this assumption that each of these features are independent of each other given s, given the sense s. So suppose there are features f1 to fn. So I was write it as orgmax s P s i is equal to 1 to n probability f i given S and this I will do for each sense and I will I will take the one that is giving you the maximum score.

Now, what is important here is what are different features that you will be choosing for a Naive Bayes method. And these features have to come from the context around the word.

So what are different things I can use in the context? I can use water different part of speech that are being used, what are different words and so on. Let us see for this task what are the important features. So this we have already seen.

(Refer Slide Time: 24:58)



Training for Naïve Bayes

- 'f' is a feature vector consisting of:
    - POS of w
    - Semantic and Syntactic features of w
    - Collocation vector (set of words around it) → next word (+1), +2, -1, -2 and their POS's
    - Co-occurrence vector
- Set parameters of Naïve Bayes using maximum likelihood estimation (MLE) from training data

$$P(s_i) = \frac{count(s_i, w_j)}{count(w_j)}$$

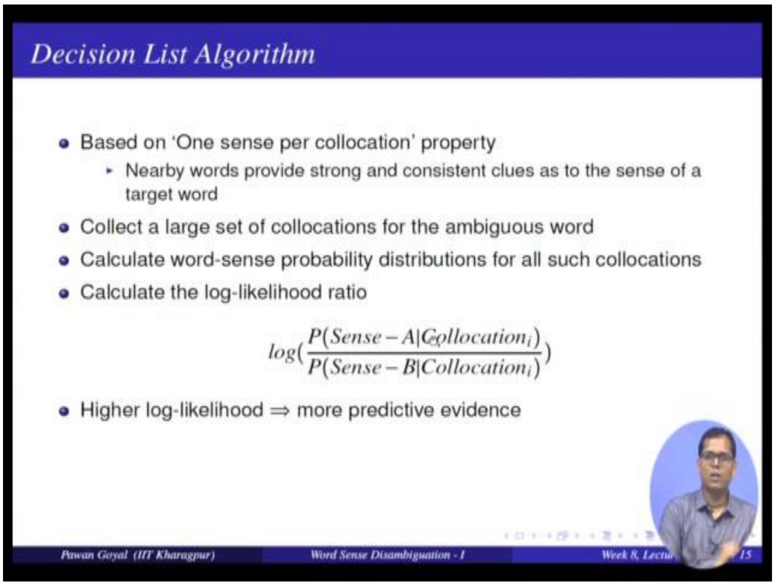
$$P(f_j|s_i) = \frac{count(f_j, s_i)}{count(s_i)}$$

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lectu 15

Some for this task, I can use the part of speech of surrounding words, and what are different semantic and syntactic features. I can use co-occurrence vector that is what are the different other words this sense is used with. And then I can use this collocation vector that is water in general the next word, next-next word, previous word, previous to previous word and their part of speech tags, when the word is used in this particular sense. So I will construct this this side of feature vector by using this all these examples. And the 2 parameters of a model that is the prior probability of the sense and the probability of features given the sense can be computed from my corpus by using simply maximum likelihood estimate.

So how do we compute probability of the sense? S i, number of times the sensitive divide by number of times this word is used. Improbability of a particular feature to the senses sense number of times this feature is observe with the sense divide by number of times all other features are or number of times senses used. So these are simple formula for mle and you use that and plug into the algorithm and you have your Naive Bayes models 30 for use. And you can use some other approaches like decision list algorithm. So what is the idea?

(Refer Slide Time: 26:27)



Decision List Algorithm

- Based on 'One sense per collocation' property
  - Nearby words provide strong and consistent clues as to the sense of a target word
- Collect a large set of collocations for the ambiguous word
- Calculate word-sense probability distributions for all such collocations
- Calculate the log-likelihood ratio

$$log(\frac{P(Sense-A|Collocation_i)}{P(Sense-B|Collocation_i)})$$

- Higher log-likelihood ⇒ more predictive evidence

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lectu 15

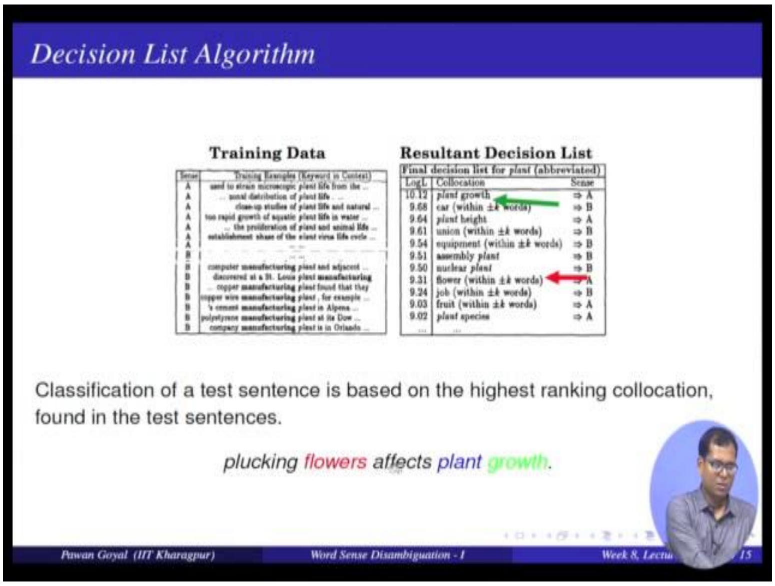
So you will use this hypothesis of one sense per collocation. Now what is this one sense for collocation hypothesis? For a given word idea is that with the particular collocation it is use only in one sense. For example, take the word like bat. We saw there are 2 2 different sense. One is like it creates cricket bat and what another it like a flying mammal. Now suppose you take this collocation cricket bat. So the word cricket covering before bat. With this collocation the word bat will always use in one sense only not the other sense. This is the idea. Can I get with the collocations to find out with this collocation this word to be with all in this sense and with the other collocations it will be used in another other senses. And this is called the one collocation per sense property. Sorry one sense per collocation property.

So how do I start about getting these set of collocations for an ambiguous word? So initially I can try out all the possible set of nearby words that that occurs with this word more often. And once I have done that for each such collocation I can compute what is the probability of a particular sense being used for this collocation with respect to the other sense. Now take the simple case where the word has only 2 sense: sense a and sense b. So for a given collocation I will find out what is the probability of sense a given the collocation, and what is the probability of sense b given this collocation.

Now, question is once I found both these numbers, what will be a function that will tell me how good this collocation is. Now this collocation would be good if one of these probabilities is high another probability quite low. So collocation indicates sense a if this probability divide by this probability is high and to indicate sense b if the inverse of that is high. And this simple measure we used to come find out what collocations for a word are more important than others.

So I come this and I take a log of this and this is called a log likelihood ratio. Log of probability of sense a given the collocation I divide by probability of sense b k given collocation i. And then hire log likelihood means more evidence. So what I will do? For different collocations that I have extended for the word these can be some simple words that occur a lot with this word in different context. I will compute this log likelihood this course. And then I am I will arrange them in their decreasing order and this will give me the decision list.

(Refer Slide Time: 29:43)



Decision List Algorithm

Training Data

Sense	Training Examples (Keyword in Context)
A	used to strain microscopic plant life from the ...
A	... zonal distribution of plant life . ...
A	close-up studies of plant life and natural ...
A	too rapid growth of aquatic plant life in water ...
A	... the proliferation of plant and animal life ...
A	establishment phase of the plant virus life cycle ...
A	...
B	...
B	computer manufacturing plant and adjacent ...
B	discovered at a St. Louis plant manufacturing
B	... copper manufacturing plant found that they
B	copper wire manufacturing plant , for example ...
B	's cement manufacturing plant in Alpena ...
B	polystyrene manufacturing plant at its Dow ...
B	company manufacturing plant is in Orlando ...

Resultant Decision List

Final decision list for *plant* (abbreviated)

LogL	Collocation	Sense
10.12	*plant* growth	⇒ A
9.68	car (within ±*k* words)	⇒ B
9.64	*plant* height	⇒ A
9.61	union (within ±*k* words)	⇒ B
9.54	equipment (within ±*k* words)	⇒ B
9.51	assembly *plant*	⇒ B
9.50	nuclear *plant*	⇒ B
9.31	flower (within ±*k* words)	⇒ A
9.24	job (within ±*k* words)	⇒ B
9.03	fruit (within ±*k* words)	⇒ A
9.02	*plant* species	⇒ A
...	...	

Classification of a test sentence is based on the highest ranking collocation, found in the test sentences.

plucking flowers affects plant growth.

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lectu 15

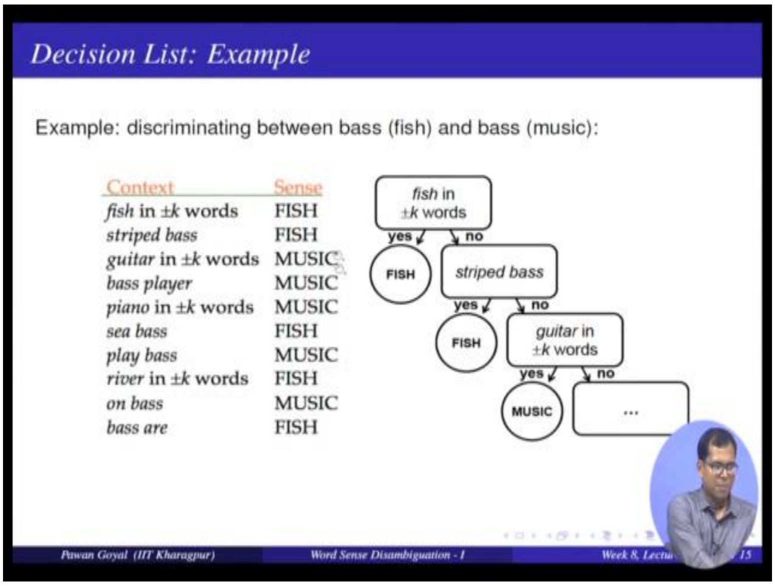
So here is an example. Suppose you have some training data and here the ambiguous word is plant, near trying to extra various collocations that can help me to find out whether the sense used is a, a is in the sense of plant life and with the sense is b in the sense of manufacturing.

So suppose I run my algorithm and I find out that the collocation plant growth it is having a very high likelihood that is 10.12. And this for the sense a, what does that mean? Log of probability sense a given plant growth divides by probability sense be given plant growth will become 10.12. And this is the highest among all the collocations. This comes out on top n a c sense is a. Suppose, similarly the second collocation is if the word car occurs in plus minus k words around this ambiguous word plant. Then the sense would be b and this as a log likelihood of 9.68 and so on. And with each likelihood I have a sense now.

Once I have that, I can use it directly as my decision list classifier. So what I will do at runtime whenever I am given the sentence I find out if this collocation is present plant growth. If this is present senses a if not whether the word car occurs in plus minus k words around plant if so senses being. If not if plant height occurs senses a and so on. So when I get a sentence at one time like plucking flowers affects plant growth, I will take this sentence and run it through this decision list classifier. And accordingly wherever I find a match I immediately provide the sense and I stop.

Now, one thing you have to be careful here with these numbers. So remember the formula we have a probable determine the denominator and it can also be 0 in some cases. So you might have to use some sort of smoothing you can use add one a smoothing or some other smoothing method for that. So once you do that you can come up with this decision list classifier and at runtime given a sentence you can easily compute what is a sense that should be used here.

(Refer Slide Time: 32:08)



Decision List: Example

Example: discriminating between bass (fish) and bass (music):

Context	Sense
fish in ±*k* words	FISH
striped bass	FISH
guitar in ±*k* words	MUSIC
bass player	MUSIC
piano in ±*k* words	MUSIC
sea bass	FISH
play bass	MUSIC
river in ±*k* words	FISH
on bass	MUSIC
bass are	FISH

Pawan Goyal (IIT Kharagpur) Word Sense Disambiguation - I Week 8, Lectu 15

Here is another example of decision list classifier. Like you are discriminating with me bass and base - the fish and music sense. So it can be something like that. Suppose this is how your different collocations are ordered as per your log likelihood. So we make a finishing tree if the word fish occurs in plus minus k words if yes sense is fish if no, if the collocation stripe bass occurs. If yes fish if no if the word guitar occurs is directly coming from the ordered list of log likelihood. And this you can run through a new sentence to find out the appropriate sense of the word bus.

So we saw some of the algorithms that that that by using knowledge based approaches and machine learning approaches. Now in the next lecture we will also talk to talk about some other approaches that are either semi supervisor unsupervised, but in general there are many different algorithms that you can apply on this task we only providing you very brief ideas on some of those.

Thank you so much. So I will see you in the next lecture.