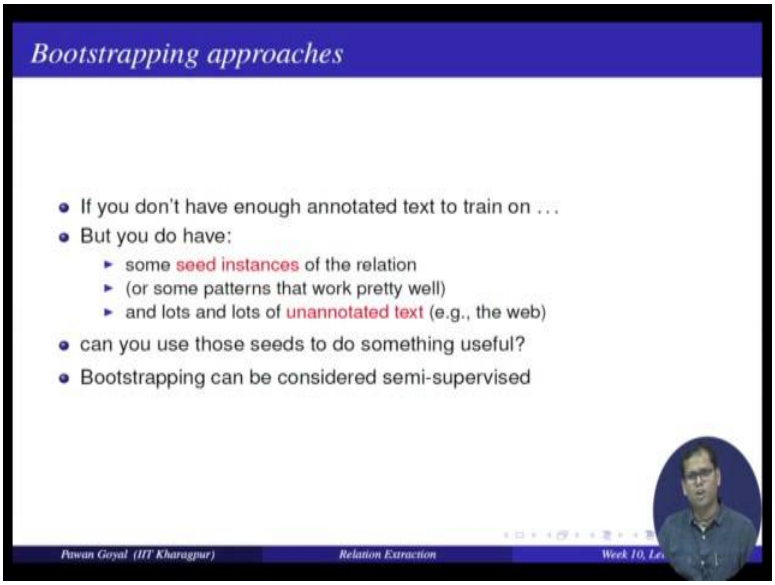
Natural Language Processing
Prof. Pawan Goyal
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 49
Relation Extraction

Welcome back for the 4th lecture of this week. So, we started talking about information extraction and there we discussed about the particular problem of relation extraction and we discussed some so, one approach that is using hand built patterns. So, we can build some patterns manually and we saw that by using that we can extract certain relations between entities. And there were some limitations with that. So, we will now see some other approaches apart from hand built patterns.

(Refer Slide Time: 00:46)



Starting with the bootstrapping approaches, so we had talked about bootstrapping approaches in one of the earlier topic that is on word sense disambiguation, but let us see how do we apply these approaches for the task of relation extraction.

Here you will use these approaches only if you do not have a lot of annotated data because if you have some annotations already available then you can use some supervisor approaches that might give you better results, but suppose you do not have annotated data; that means, data where it is labeled entity 1 is related to entity 2 by a

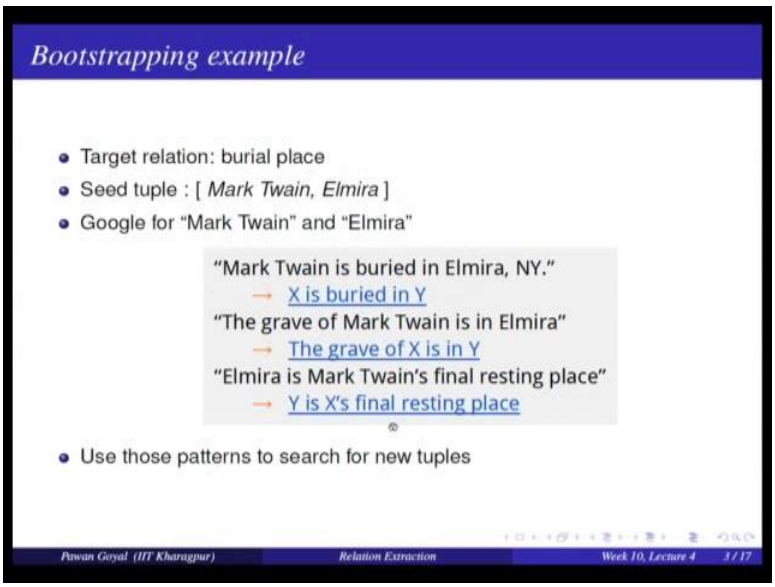
particular relation R, if you do not have such data you will use your bootstrapping approach.

You do not have enough data, but what you have r some seed instances of the relation and that is very easy. Now what do I mean by seed instances? So, remember we were talking about hyponyms or meronyms, see all you know some seed relation. So, you know basement and buildings are connected by the meronym relation, car and vehicles are connected by the hyponym relation.

You know some seed instances then what else do you need? Or you might have some patterns that you know and lots and lots of unannotated data; that means, you should have a lot of corpus where you have a lot of text data it may be an unannotated, it is a simple text data. So, idea is using some seed patterns, running some idea over this whole data you can try to bootstrap your approach for relation extraction.

The questions here are, so suppose so here you have some seed instances. So, how do you use them for doing something meaningful or so, that you can extract where is other entities? So, you have some seed instances how do you use that for extracting more such examples and so, this can be consider some sort of a semi supervised approach. So, what do we do here? So, let us take some simple example.

(Refer Slide Time: 02:55)



Here suppose I have my relation is burial place. So, X is buried in place Y, I want to find out all such entities that are connected by this relation.

How do I go about it? Firstly, I need to see if I have some seed instances that is some entities that are connected by this relation suppose I have an instance that is Mark Twain is and the burial place is Elmira, New York. So, I have this seed tuple, now how do I start? So, remember what did we need? We need some seed instances and a lot of corpus data.

Now you can use the intuition that how are you finding out the patterns in the case of hand built patterns. So, again we were starting with some hand with some seed instance. So, like basement building and we were trying to go to the corpus and seeing wherever these 2 words occur what is the pattern in which they occurring, same thing you can do here. So, you know I have 2 entities Elmira and Mark Twain. Now you can search your whole corpus to find out where all these 2 entities occurred together any single sentence or in some proximity.

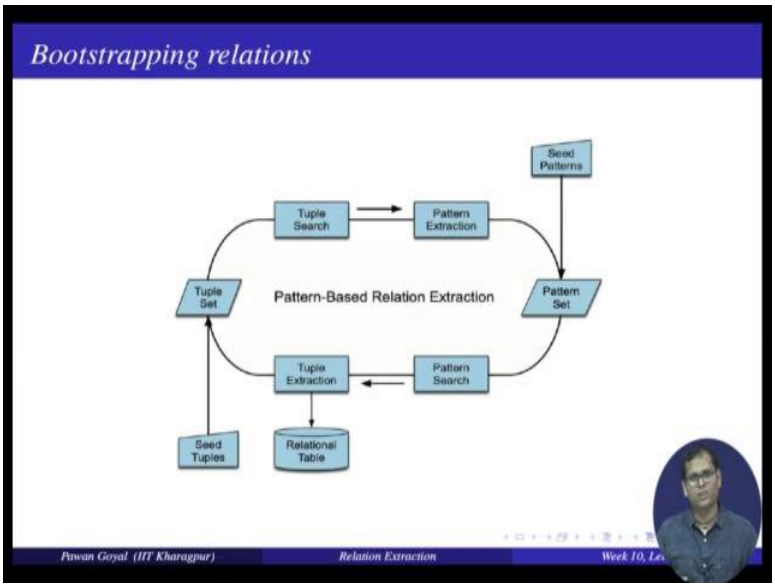
Now, wherever they occur you try to find out in what pattern do they connect to each other and these patterns you can generalize as your gen generic patterns without having to manually go through each of these? So, something like so, I have these and it is Mark Twain, Elmira and maybe I can google this google this google this entities Mark Twain Elmira together and let us see, what do I; what are the kind of sentences do I get and suppose I get a sentences like this, excuse me, Mark Twain is buried in Elmira New York, this is the sentence and I know what am I entities here Mark Twain is X Elmira is Y. So, I can immediately from a pattern X is buried in Y in this become my pattern.

The next sentence the grave of Mark Twain is in Elmira. So, I can have this pattern the grave of X is in Y similarly Elmira is mark twins final resting place. So, I have this pattern now Y is X S final resting place and so on. So, we are getting this sentence is from the sentence you are extracting over you are entities X and Y and you are building patterns in terms of X and Y. So, what you saw here just by using 1 seed instance and a lot of unannotated data you can find out some patterns.

Now, once you have patterns, what you will do? You will use these patterns to find out more and more such entity pairs that are connected by this pattern and that will enhance your seed tuples or seed instances then you can again use this seed instances to again

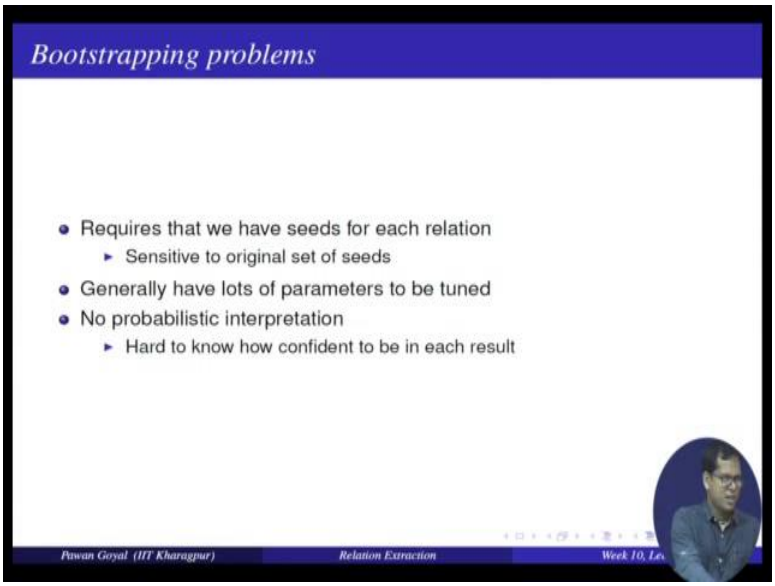
such the corpus find out more such patterns and this can be done in an iterative manner until there is some convergence going on or you are seeing that there is not helping much.

(Refer Slide Time: 06:06)



Now, you have these patterns and you use these to search for new tuples. So, describe described by the simple flow chart that is you are starting with this some sort of seed tuples like Mark Twain Elmira then you are searching with tuples in the corpus and you are finding various patterns and this becomes your pattern set you might also have some seed patterns already, but you are getting some patterns. Now using this patterns you are searching the corpus and finding more tuples and I putting them in your relational table and that is going to your tuple set, now using your tuple set you can search these and find out more patterns and this can keep on going in many many iterations and this is in a very nice approach you can see that you only need data that is freely available everywhere and you need some very few seed instances and you can apply this algorithm and this does not require to do everything manually.

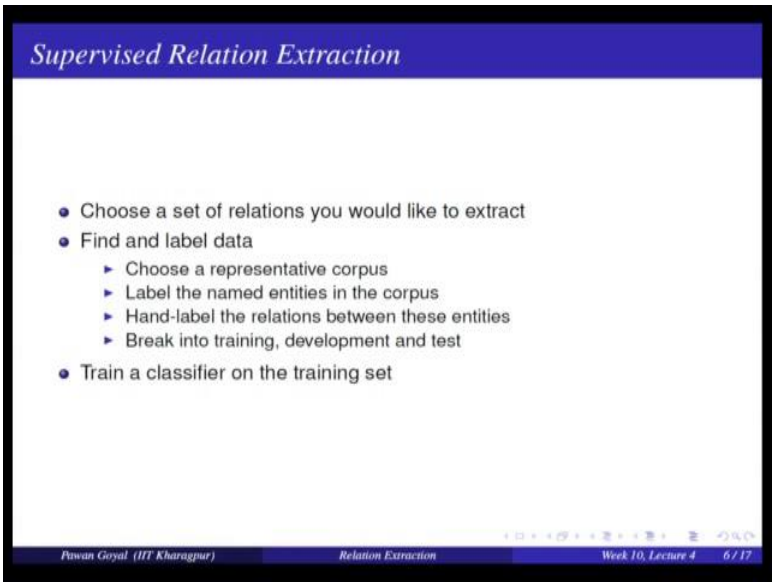
(Refer Slide Time: 07:01)



Yes, but there are some problems with that approach the problems. For example, are that the seed instance with which we start should be a good instance such that there are many occurrences of this seed in the corpus if the seed pair does not arrange the corpus, you will not be able to extract many relations many patterns from this. So, this is one problem with this approach. So, this is sensitive to the original set of seeds that you use for your algorithm and in general there can be many parameters to be tuned, for example, how many top patterns will I take from my set? How many iterations I will go through and how many times I will send the same pattern for my first search? There can be many a many parameters that you have to fix and yeah there is no such probabilistic interpretation. So, it is difficult to know how confident you are in each pattern or each tuple that you are finding by this approach.

Here some sort of problems with this approach, but it is a nice approach if you use to want to get it done without building some sort of machine learning method also, so on, you can use this approach very for some simple and easy results, but we will see if you want to build a more over system that what kind of approaches you can you can use.

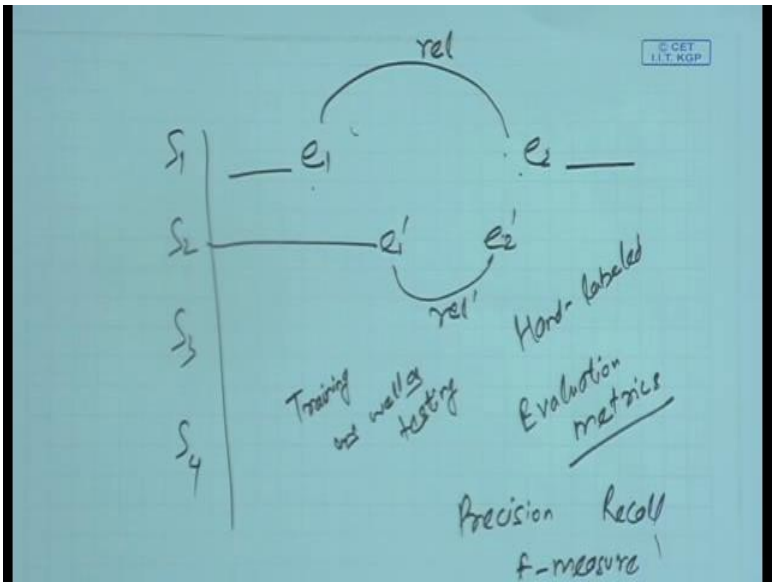
(Refer Slide Time: 08:21)



For that we can talk about some supervised approaches for relation extraction. So, what is the idea? So, first you will define, what are the kinds of relations you want to extract? So, relations can be many like I want extract family relations. So, parent of wife of husband of and so on, you can extract some organization relation, this is an employee of and subsidiary of and so on. So, we will define a set of relations.

Now, for each relation, you will find data and label the data. So, they should be some manual labeling involved somebody has to label the data that in this sentence these entities are connected by this relation. So, you will choose a representative corpus where you think that there can be some instances of this relation now you will label the named entities in the corpus and hand label the relations between the entities.

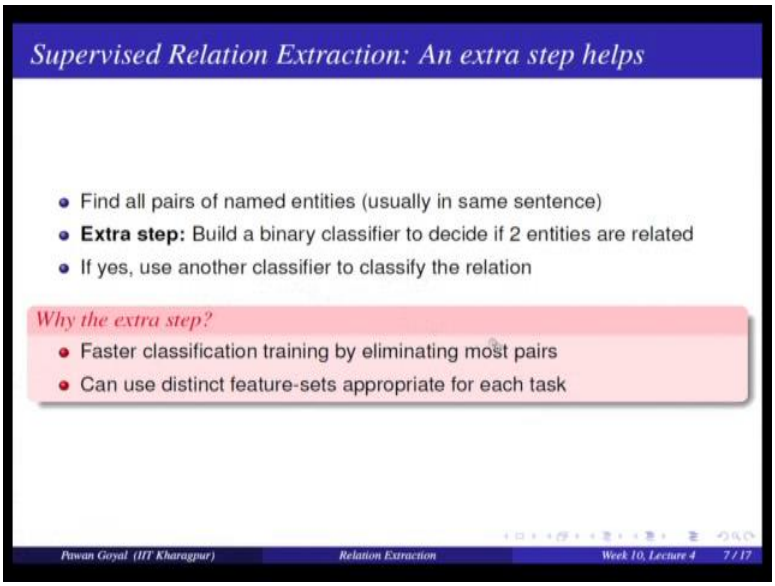
(Refer Slide Time: 09:25)



What will happen? You have a corpus you will find out sentence S 1, S 2, S 3, S 4 and you say in the sentence, this entity 1, this entity 2 and you know what is the relation between them, similarly here you find there is 1 entity; 1 prime entity, 2 prime and there is some relation prime here and this has to be hand labeled why do you need hand labeling. So, once you have these hand labels they are like your; so this is like your bold standard that you can you as your training as well as testing data.

We will train your system using in this sentence, if these are the entities there is a relation between them. So, in a new sentence suppose 2 entities are there is there is the relation between them. So, this can be some machine learning model that you can built by using this gold standard and then you will yeah break into training development in text that is the usual practice in machine learning and then you will train a classifier on the training set.

(Refer Slide Time: 10:39)



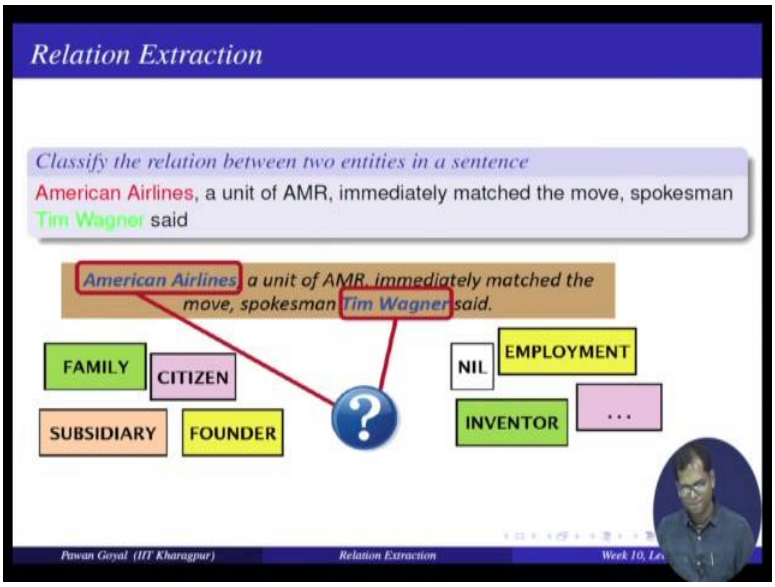
Now so here it might be one problem while you are using this approach in general, they can be 100s of relations and so you need to; so when in you are given a sentence between 2 entities take they can be many relations, but they may not be any relation at all, it might happen that 2 entities are occurring in a sentence, but they are not connected by any relation (Refer Time: 11:04) there is no relation as such.

What you might do is to have the first step that says in a sentence, I know what are all the entities and this first step tells me which 2 entities are connected and which 2 are not connected and once you have the output of this step, you know these 2 entities are connected then you run your additional classifier to find out, what is the relation between them among all the 100s of relations that you have.

This issue; this is seemed to be working in this in this area of relation extraction that first you find out what entities are connected second what is relation between them. So, it is a 2 stage approach. So, you find all pairs of named entities in a sentence and the extra step can build a simple binary classifier. So, that is says yes or no and it decides whether 2 entities are related or not and if you find an answer, yes, to this is step use another classifier to find out what is the relation between these 2 entities and why will that help because in the first step build itself you will be able to eliminate a lot of extra pairs that are not involved in any relation. So, we will not bother about those you will only bother about those entities that are probably connected by some relation.

Other advantage could be you can think of the idea sort of features that you will used for the first step that is finding out if there is a relation or not and the second step that is if this relation what is that relation you can think of very set of features that you can imply both for both of these steps.

(Refer Slide Time: 12:46)



Here is one visualization of what this will look like. So, suppose you have this sentence American Airlines a unit of AMR immediately matched the move a spokesman Tim Wagner said. So, now, suppose by using the first step you found out that American Airlines and Tim Wagner are connected by a relation. So, now, you want to find out, what is the relation? So, here you have a sentence, 2 entities you need to find out, what is the relation among all these possibilities is family relation, citizen relation, subsidiary relation, founder relation and so on. So, lots of relations are there you want to find out what is the relation between these 2 entities.

Now, how do you solve this problem? How would you solve this problem? So, you will have a lot of labeled data when you know these are the entities here and this is the relation between them. So, in classification, what we do from this labeled data? We try to abstract over some sort of features. So, we will say so, these are the features that I see in this sentence and these features indicate relation 1, other sentence I am seeing this kind of features that indicating relation 2 and so on. So, this is what I will have from my training data. Now attached data again, I will try to find out what are the features and

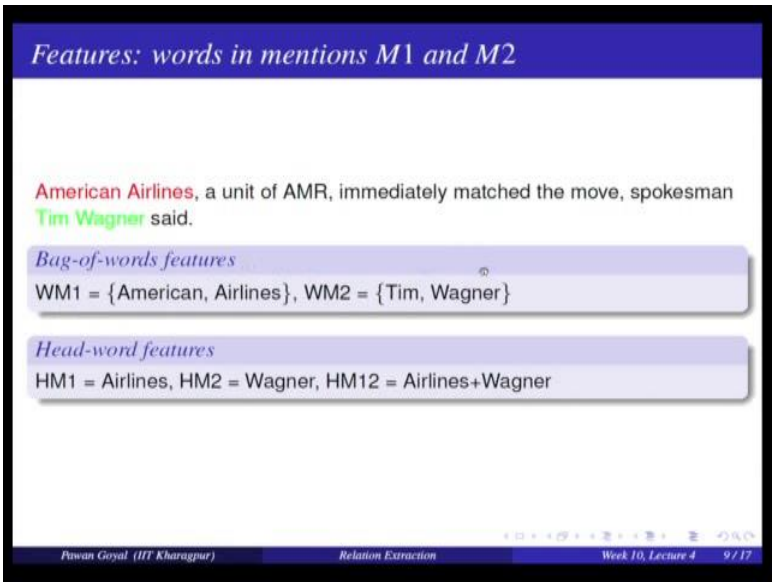
using these features I will try to match with one of this previous examples I have seen in training data, this is simple illustration, but it is generally more complex than that, but this is the basic idea.

So, the whole effort goes in deciding, what should be my ideal features by which I can represent all my data points. So, how do I say, these 2 are connected by this relation? What are the different things in the surrounding, in the context about these entities that I should be using to make this decision? And that is your task of each engineering find out what are the features that will help you in this task.

In most of the NLP application, this is one of the main challenges that for this task find out, what are the appropriate sets of features I can use. So, we will see some examples at what? So here you can use all the different concepts that are covered in this course. So, it is starting from simple language models part of speech tags dependency parse synthetic parse everything, you can use to do find out to define, what are your features in this task and you will see a lot of examples here and this is one. So, if you want to build your own system you might have to start thinking in terms of what are the important insights from data that I here use as in the form of my features?

Remember, features are something that you think can help me discriminate between various relations here. So, it can help me, tell me tell if this is a family relation versus if it is a citizen relation, what are the different things that can help are these various words that occur in the context are these part of speech tags and or this is a something else. So, this you can abstract in terms of here features.

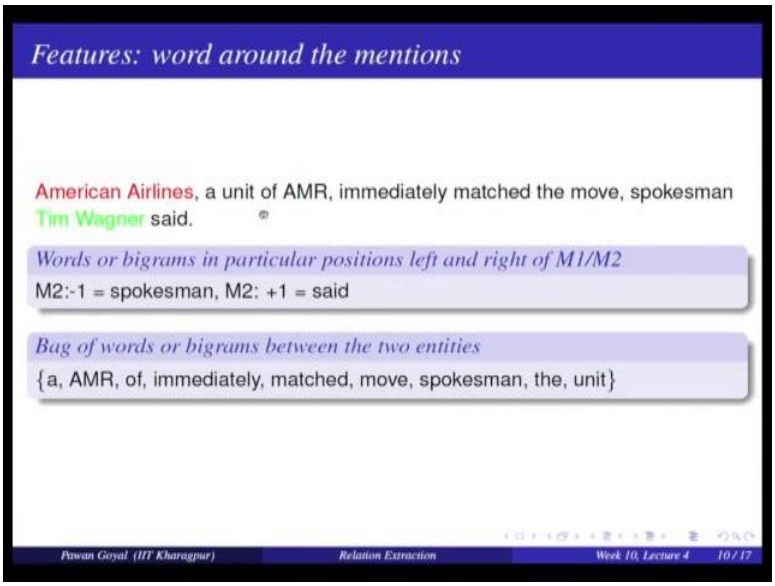
(Refer Slide Time: 15:54)



Let us see what are the kinds of features, we can use for this task. So, I have the sentence American Airlines etcetera and my initial features could be what are the words in my mention; M 1 and M 2? So, M 1 is American Airlines M 2 is Tim Wagner. So, what are the words that I used in these 2 mentions? So, feature here can be bag of words features. So, mention 1 uses word like American Airlines and mention 2 uses words like Tim Wagner. So, these are simple features I can also use what are the headwords of these 2 mentions the headword mention of mention 1 is airlines and for mention 2, it is Wagner and you can also see, what is the headword mention of 1 plus 2 airlines plus Wagner?

Why you are using this headword kind of features? So, here you are having American Airlines, but suppose there is something like Indian airlines or some other airlines. So, by using the headwords using capture E 1, a new word that has the same headword, but the initial word was different it can be captured by using headwords same with Tim Wagner. So, you are capturing the surname here by headword, but suppose if someone else has a surname Wagner. So, it can also be used.

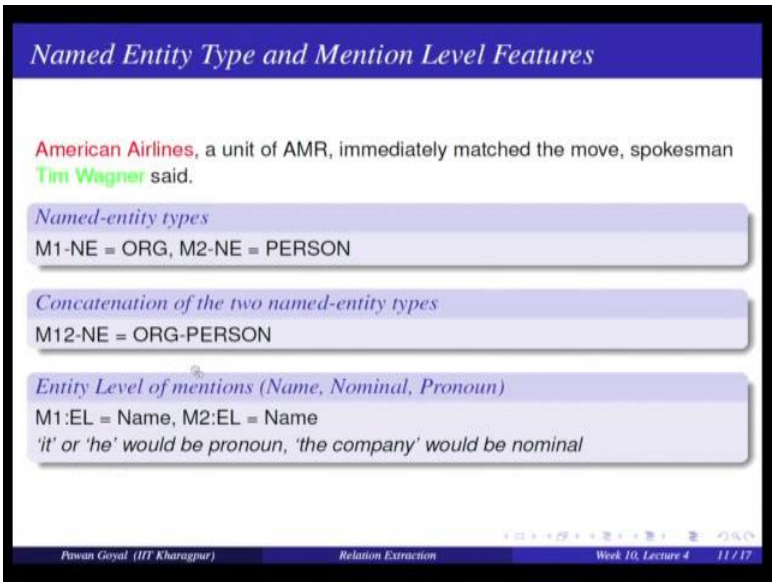
(Refer Slide Time: 17:11)



Then I can use, what are the words that are coming around the mentions? So, that is what are the words are coming before American Airlines after Tim Wagner and what are the words in between? So, what can be my features? So, words or bigrams in particular positions left and right of M 1, M 2 like what is the word before M 2. So, it is a spokesman, what is the word next to M 2? That is said, so what you are abstracting here? I have an entity before which I have a word spokesman and next word is said. So, the new context whenever I see, what is spokesman before, what said afterwards, it might indicate that there might be this relation, a spokesman X said it might be a good indicator of this relation.

You can also use the back of words or bigrams between the 2 entities that is what are the different kind of words that occur between the 2 entities here? So, we will say. So, words like AMR immediately matched a spokesman unit etcetera they are all occurring between the 2 entities and they all go as your features.

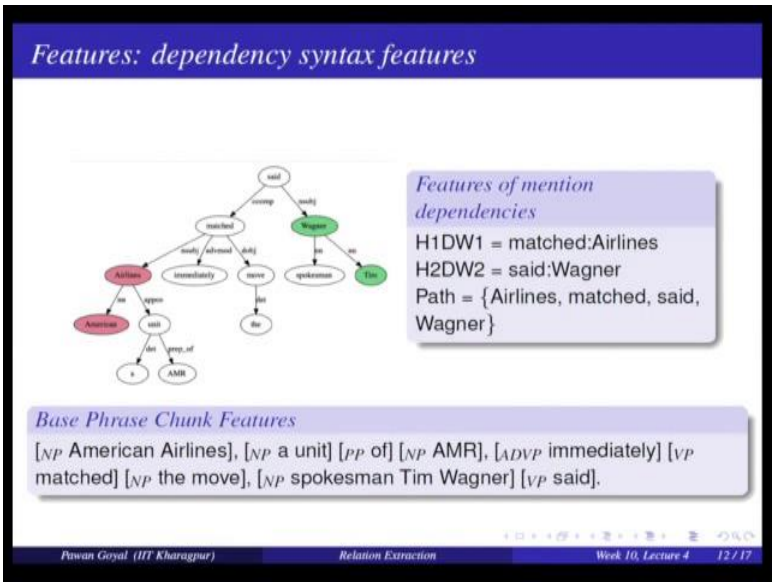
(Refer Slide Time: 18:18)



You can have named entity type and mention type features. So, for example, what is the named entity tag for the mention one? So, it is like American Airlines organization. So, this you can get by using various named entity recognition tools you can run in any area and you can find out what are the various named entities. So, it is say mention 1 is in organization.

Similarly, mention 2 is a person. So, this can be nice feature that can help you, this is an organization, this is a person. So, what can be a relation between them and yeah it can be together also, what is the named entity for 1 in 2 together organization person then you can also find out entity levels of mentioned is the name nominal or pronoun. So, here first one is a name second one is also a name, but suppose in the sentence you have it he etcetera. So, we can call it as a pronoun on the other hand if you have a word like the company you will call it as a nominal. So, all these can also be your features.

(Refer Slide Time: 19:25)

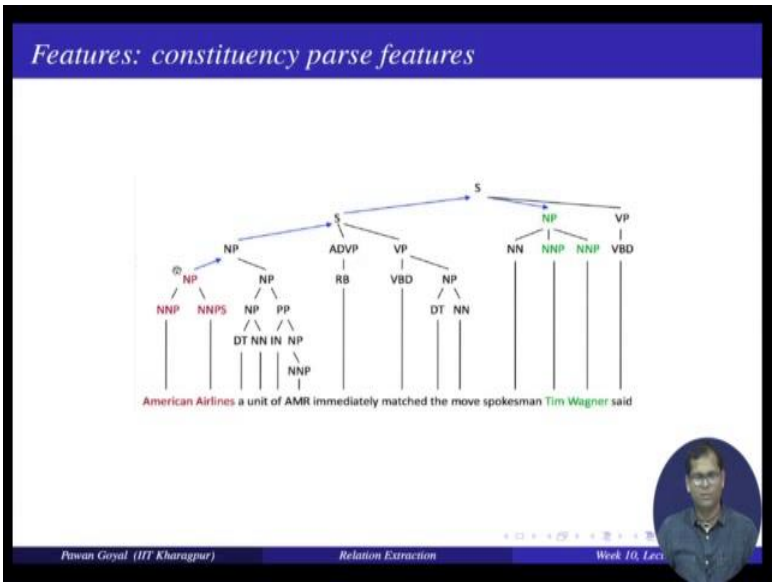


Now, suppose you want to use the dependency between them. So, you will see when you convert the sentence to a dependency graph, what is the connection between the 2 entities? What are the different branches in that in the tree by which they are connected? So, suppose you find this dependency graph, so we will say ok. So, they are connected by this path matched, said and you, are saying going to Wagner airlines matched, said, Wagner. So, now, you will try to use certain features based on this path also. So, it can be maybe what are the words that occurring in this path or what is the complete path altogether?

Here, what is the headword of the dependence; the dependency headword for word 1. So, you had airlines matched airlines is the dependency for the word one for headword 2. So, we have the dependency said and Wagner this can be a feature immediate dependency feature matched airlines said Wagner then what is the path airlines matched said Wagner and you can also think of some other features what is the label they are at in that dependency graph how many different words they are connected to and so on.

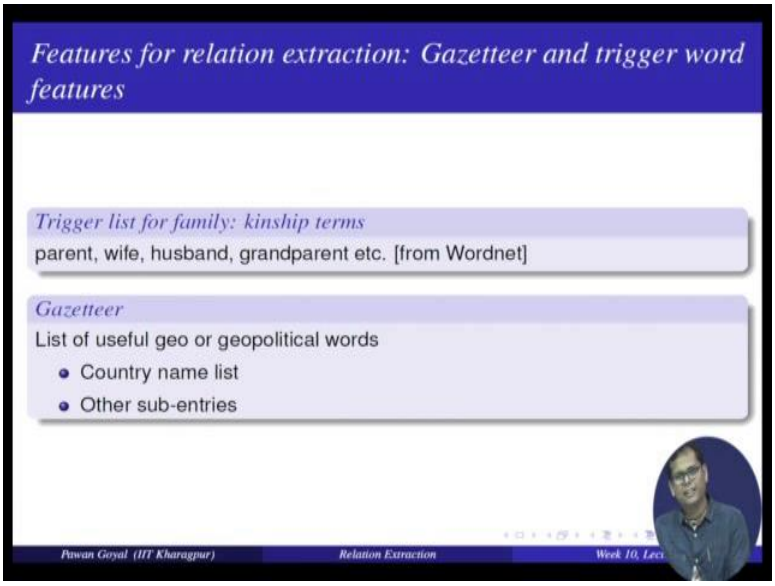
And then you can also do chunking and is use used features like if you chunk them you will find American Airlines a unit of AMR etcetera and your feature could be what is the chunk in which it participates what is the next chunk after this and so on.

(Refer Slide Time: 20:58)



You can also use the constituency parse feature. So, you have the 2 words here, American Airlines and Tim Wagner and this is the party of the sentence. So, you can use the path from here for noun phrase to this particular noun phrase that connects Tim Wagner. So, what is the path here going to an noun phrase to a sentence to a sentence to a noun phrase this path can be helpful again here you can use what is the label in the tree they are at and what is the sibling and so on. So, these can be your various features.

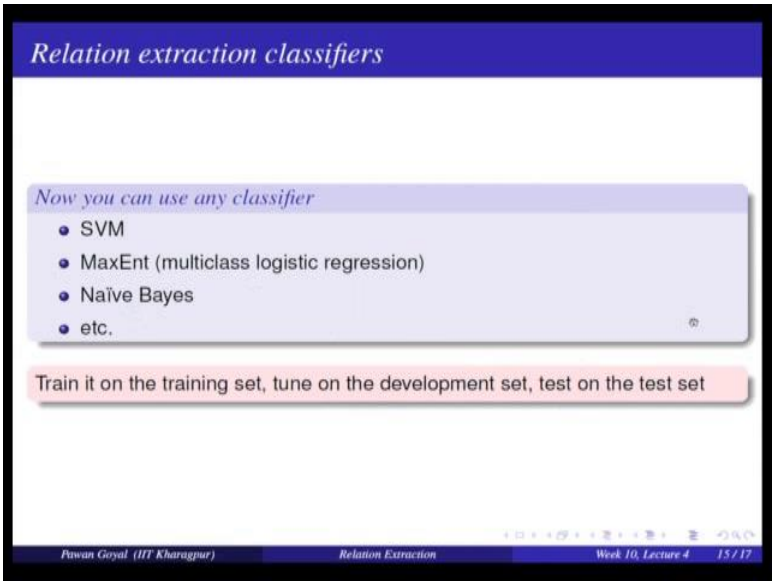
(Refer Slide Time: 21:32)



Then they can be some sort of features that you can obtain from various gazetteer and different terms kinship terms. So, if your relations contain mother of and parents of an all that. So, you can use some kinship terms. So, like parents wife husband grandparent etcetera and this can be obtained from various electrical sources like Wordnet also and then you can use various gazetteers like, what are the country name lists? So, you can see.

The next word after the entity 1 is a name of the country or the previous word after the entity 2, any of a country, similarly for names of very famous celebrities or persons all these can be used as your features. So, this is the idea that that you have this task you have to find out the relations and you should be able to use whatever sort of features you think will help in this task. So, we are seeing we are using a lot of different sort of features and so and they are using all the different concepts and topics that you have seen in the basics of this course.

(Refer Slide Time: 22:44)

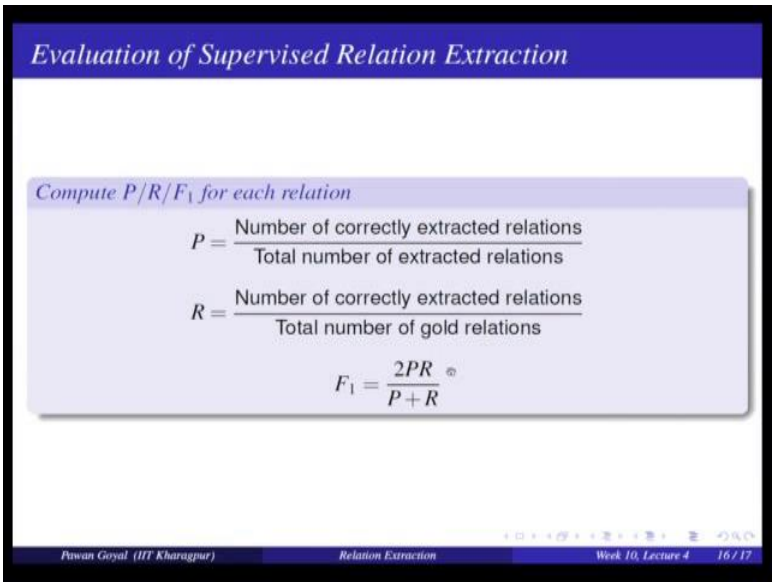


You can have country name list others of entries etcetera. So, now, once you have identified what are your features, so what will happen if you are training data, each instance you can convert in a feature vector. So, we know what are the different features that are that are involved in this particular sense and then you have you can use various classifiers to built to find out given a new sentence how do I find out if the 2 entities have a relation and what is the relation and then you can you can use multiple classifiers

like naive Bayes classifier or SVM or MaxEnt etcetera and this is the rule always you train on the training set tune on development set and test from the test sets.

You should never use your test set for building your features or whatever patterns. So, you should never use it as set you should be kept separate. So, you will only tune on your development set if you your training set you run your classifier and initially test on the development set if it does not work keep on improving and once you are satisfied then the on the test set and find the accuracy and you might also have to compare with others if you want to find out if you are doing something better than what people have already done.

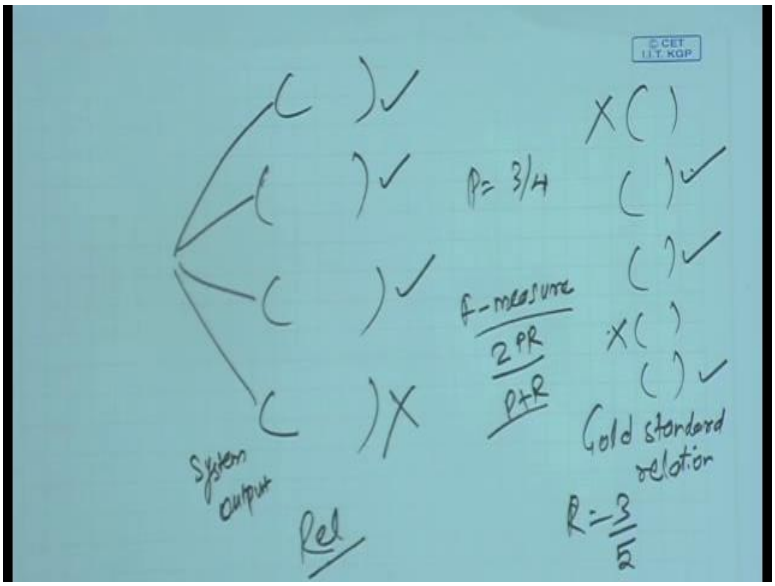
(Refer Slide Time: 24:01)



So, now how do I, so once you have done this classifier, you will get your; so classifier will tag in this sentence these 2 entities are connected by this relation and so on. Now how do you find out how good your system is performing and how we compare with the system. So, for that we will talk about what are the various evaluation measures or evaluation metrics matrix, how do I evaluate it? So, in standard that evaluation metrics are, what is the precision? What is the recall and what is the F major?

So, how do I define these? What is the precision recall etcetera? So, here are the simple definitions. So, precision is so for suppose, I am doing it for each relation separately. So, precision would be what are the number of correctly classified or extracted relations divided by total number of extracted relations?

(Refer Slide Time: 25:03)



That is suppose my system is giving, so these entities these pair of entities are connected by a relation R, this is the output of my system.

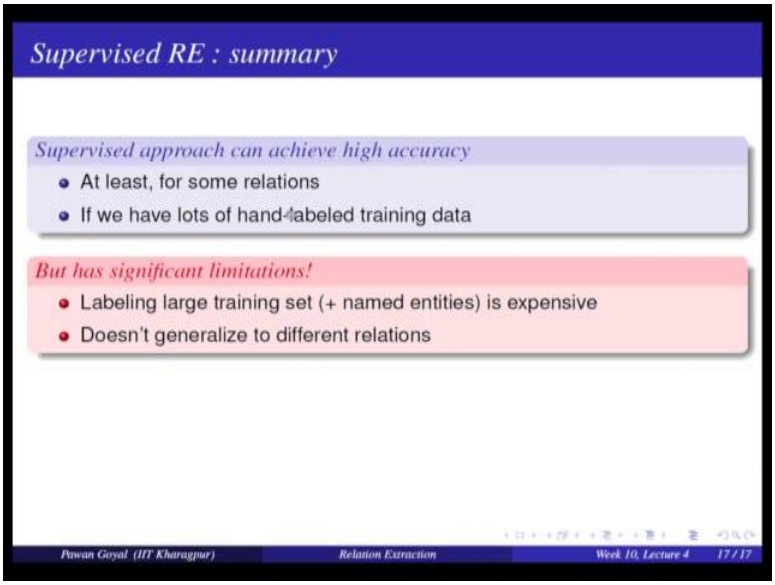
Now, from a gold standard, suppose I know that this and this is correct and this is incorrect, system will give 4 output; 3 are correct. So, precision here will be 3 by 4. So, this is one very important metric that my system should have a high precision; it should be whatever relational predicting should be correct, what is the other criteria? So, the criteria is recall is what are the number of correctly extracted relations, but my system identified divided by total number of gold relations, now it is important to find the see the distinction between that 2.

So, what we will do in re recall? So, the system output in requires the see what is the what are the gold standard relations supposing my gold standard I had 5 entities I know these entities are connected by this relation and suppose my system has found out. So, it has found out 3 it is found out this, this and this, but my system could not find out this and this. So, what is the recall of my system my system could recall 3 out of 5 possible relation. So, recall here would be 3 by 5 and this I can do for every relation independently and I can show my system does work for this relation with this precision this recall and so on.

Now, there is also a metric where you can combine these 2 and that is called f measure and what you do is to take some sort of harmonic means. So, this is 2 P R divided by P

plus R, you will take a harmonic mean of precision recall and that is called your F measure. So, this is F measure 2 P R divided by P plus R, although there are variations where you can give different weights to precision and recall also, but this is quite accepted measure that you give a equal weightage to precision recall and find out an F 1 measure.

(Refer Slide Time: 27:24)



Now if we try to summarize the supervised relations extraction task, what we did here? So, in general, it can achieve very high accuracy for some relation and if we have lots of hand labeled training data. So, for most of the machine learning algorithms, they all depend on how much label data that you have. So, if you have lots and lots of data, they can give you better, better and better accuracy. So, that is one bottle neck also. So, you need to label lot of data to be able to get good accuracy and. So, this is the limitations here. So, labeling large training set and the entities might be very very expensive and it may not generalize to different relation. So, I have labeled for some relations, but suppose I want to now extract some new relation I cannot use this label, I need to get new labels for the new relation

Whatever I have labeled, the model will only be able to extract those relations a new relation I have to do the labeling again. So, this also does not generalize. So, we sought to approaches. in this lecture, we saw bootstrapping, you have you do not want to annotated lot of data here, some seed instance you go for that, but you can if you can

annotated lot of data then you can use a supervised approaches and they are the main problem is after labeling find out interesting features that can help with this task.

Now, in the final lecture for this week, you will see an interesting approach that blends these 2 approaches together; bootstrapping and supervised approaches and this can also generalized to many many different relations that that you have. So, this is take this approach is called distance super provision. So, in the next lecture, we will talk about that approach.

Thank you.