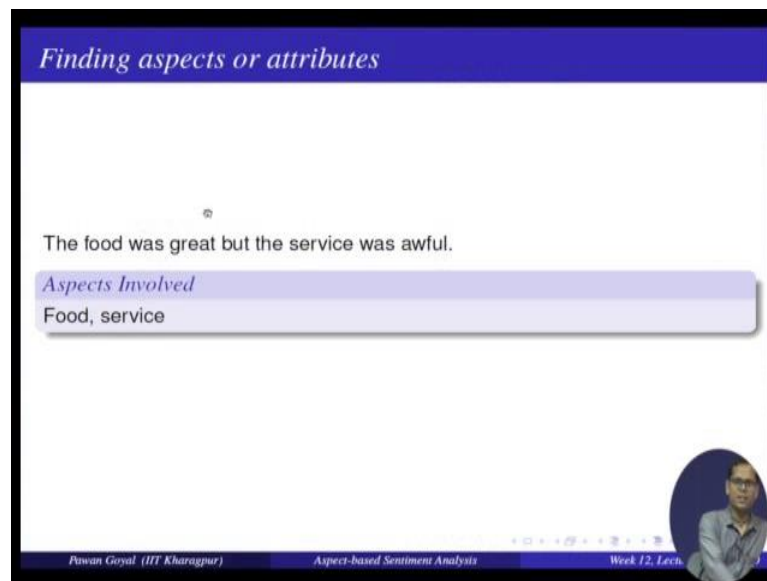


Natural Language Processing
Prof. Pawan Goyal
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 60
Aspect Based Sentiment Analysis

Hello everyone, welcome to the final lecture of this week and also this course. So we will be finishing the topic of sentiment analysis. So, we discussed a lot about what is sentiment analysis, how do you obtain sentiment lexicons, how do you learn those how do you compute with those use some linguistic intuitions and so on. In this final lecture we will discuss few things like how do you use them some for some other applications like one of the application aspect base sentiment analysis. So, what do I mean by that.

(Refer Slide Time: 00:57)



Finding aspects or attributes

The food was great but the service was awful.

Aspects Involved

Food, service

Pawan Goyal (IIT Kharagpur) Aspect-based Sentiment Analysis Week 12, Lec.

The slide features a blue header with the title 'Finding aspects or attributes'. Below the title, a sentence 'The food was great but the service was awful.' is displayed. Underneath the sentence, a light blue box contains the text 'Aspects Involved' followed by 'Food, service'. In the bottom right corner, there is a small circular portrait of Prof. Pawan Goyal. The footer of the slide includes the text 'Pawan Goyal (IIT Kharagpur)', 'Aspect-based Sentiment Analysis', and 'Week 12, Lec.'.

So many of the times when we see the reviews they are not saying only positive only negative things about the product, about the hotel, about the movie, so you might have different sentiments about different aspects of a product ok. You can say I saw I got the camera and it was light weight so I am happy with that it was the price was ok it was not too expensive I am fine with that, but the picture quality is not good. So, you are not saying completely negative about the camera and neither you are saying completely

positive about the camera, but if you can find out there are certain aspects and you can say to this aspect I am giving you this opinion to this aspect I am to this opinion that might be very very helpful.

Also so not only to get overall summary to what to what aspect this product is not doing well to what aspect is product is doing well and also try to you can use that to compare to products which product is better in terms of image quality, which product is better in terms of price, which term is better product is better in terms of say resolution and so on.

Now how do we capture that? So, again there are lot of lot of models that have been proposed for that we will see only some simple linguistic intuitions that can be applied for solving these problem, but there is a lot of literature around that. Let us take an example: I have this review the food was great, but the service was awful. So, you can see nearly 2 different polarities here the food was great and the service was awful.

Now as we have seen, sentiments have being expressed towards two different aspects see what are the aspects here, one is the food another is the service. So, can we capture that what are the different aspects that are being told in the reviews. Now how do we go about that? How do we find out for a particular domain what are the important aspects. So, again there are various algorithms. So, one very naive algorithm would be, we are taking that the sentences or reviews in a domain find out what kind of noun phrases or words occur a lot with some opinion words so; that means, a opinion word occurs just before them. That might be nice way of finding out what are the good aspects. Something's like this.

(Refer Slide Time: 03:22)

Finding aspect/attribute/target of sentiment

Frequent phrases + rules

- Find all highly frequent phrases across reviews ("fish tacos")
- Filter by rules like "occurs right after sentiment word"
- "... great fish tacos" means "fish tacos" a likely aspect

Casino	casino, buffet, pool, resort, beds
Children's Barber	haircut, job, experience, kids
Greek Restaurant	food, wine, service, appetizer, lamb
Department Store	selection, department, sales, shop, clothing

Pawan Goyal (IIT Kharagpur) Aspect-based Sentiment Analysis Week 12, Lec 6

So, I take all frequent phrases across the reviews like fish tacos and so on. Now among all that frequent phrases I find out those that occurs a lot right after some sentiment word. So the idea is that if it is in aspect sometimes it so at least some time it will occur with the sentiment word mostly directly after it is a good and good fish taco and so on it is getting directly after sentiment word.

So whenever you see like great fish taco it means fish tacos is likely an aspect and you do that for all the frequent phrases that you have found in the corpus and I guess it is a very nice and simple intuition and by doing that itself you might get a lot of interesting aspects. You see like yes no. If you do it on a casino domain you find casino buffet, pool, resort, beds; children's barber you get like haircut, job, experience, kids; Greek restaurant- food, wine, service, appetizer, lamb; department store you will get words like selection, department, sales, shop, clothing, and now all of you all of them you see look like some nice aspects. You can further prove if you want by seeing; what are the aspects that occur explosively in this domain but not in other domain.

So, that will say these are the good aspects for these domain itself. So, that way you can find out what are the important aspects, once you found also aspects it will go through the corpus see; what are the opinions for each other aspects.

(Refer Slide Time: 04:53)

Extraction of aspect-opinion pairs		
RuleID	Observations	Examples
R1	$JJ \leftarrow amod \leftarrow NP$	The camera has a good screen. (good \leftarrow amod \leftarrow screen)
R2	$NP \rightarrow nsubj \rightarrow JJ$	The flash is brilliant. (flash \leftarrow nsubj \leftarrow brilliant)
R3	$VB \rightarrow dobj \rightarrow NP$	I love the image quality. (love \rightarrow dobj \rightarrow image quality)
R4	$NP \rightarrow nsubj \rightarrow JJ,$ $JJ \in \text{implicit aspect lexicon}$	The camera is expensive. (camera \leftarrow nsubj \leftarrow expensive)

Another possible approach for doing that is give some sorts of linguistic insights in how do people connects in opinion aspect in a sentence. Whenever you talk about some aspect within opinion is there some buffer structure that is followed and there has lot of rules that can be formulated by that and they are based on the dependency pass. How are they connected in the dependency pass? If you see some examples here, there are four examples the camera has a good screen.

Now what you are seeing if you see dependency pairs there will be empty. That is a screen and they will be J J like good ok. Good will be adjective modify of a screen. Now this is a relation that can be captured from dependency pairs and suppose now you want to abstract that. Wherever a there is a J J and N P occurring in a dependency pairs such that J J is adjective modify and N P you can take this and say J J is your opinion and N P is your aspect.

And you can do that across all such occurrences that would be a nice rule and that can give you all possible aspect obedient pair that are occurring with this rule. Similarly here the flash is brilliant you are seeing the word flash occurs is as a subject of brilliant. Now you make a rule whenever N P is there and J J is it is and subject then you will take those aspect obedient pair. Similarly here I love the image quality, there is a verb and it is

object is a non phrase. Verb log is the verb that is my opinion here and image quality in N P is my aspect.

Similarly here the camera is expensive you can find out camera N P and J J is the subject of that camera is expensive. So, here what is interesting is that expensive is not directly in an aspect it might be some implicit aspect it talks about an aspect like price. So by that you are saying by expensive. So, like that you can formulate some rules there are rules already available in the literature that can give you various aspect of obedient pairs from your review corpus. Suppose you have understood your aspects and you have some sentences where you know this is aspect 1 is aspect 2 aspects 3 they are all labeled.

Now once we have that then your task becomes very easy you can use a supervised classifier. What will you do? It will learn classify to classify whether the sentence talks about one of these aspects. It will learn from all these aspects the labeled sentences given a new sentence it will try to classify it and one of the aspects and, so that might be a nice approach.

(Refer Slide Time: 07:42)

Finding aspect/attribute/target of sentiment

If the aspects are well understood, use supervised classification.

Rooms (3/5 stars, 41 comments)

- (+) The room was clean and everything worked fine – even the water pressure ...
- (+) We went because of the free room and was pleasantly pleased ...
- (-) ...the worst hotel I had ever stayed at ...

Service (3/5 stars, 31 comments)

- (+) Upon checking out another couple was checking early due to a problem ...
- (+) Every single hotel staff member treated us great and answered every ...
- (-) The food is cold and the service gives new meaning to SLOW.

Dining (3/5 stars, 18 comments)

- (+) our favorite place to stay in biloxi, the food is great also the service ...
- (+) Offer of free buffet for joining the Play

Pawan Goyal (IIT Kharagpur) Aspect-based Sentiment Analysis Week 12, Lecture 3 5 / 10

So, like here, you are having different sentences one some sentences about rooms. So, the room was clean and everything work fine, we went because of the free room the worst

hotel I had ever stayed at, even a stay shocking work room. Then something about service, upon checking out another couple were checking early due to a problem and everything single every single hotel staff treated us great about service, the food is cold and the service gives new meaning to slow again this is negative sentiment. But they are all occurring in their aspect of service.

Similarly what you will do, you will take different aspects and label difference sentences as per these aspects and then you use any supervised classifier like naive bayes or (Refer Time: 08:22) integration whatever and given a new sentence find out what is the aspect involved.

Now coming to some of the problems that you might face while doing that; so many a times we tend to think that a word will have only a particular sentiment scores either it is positive or it is negative, but this is not always the case you might have to build opinion lexicon corresponding to different aspects. So for one aspect this a word might have a positive meaning for another aspect it might have a negative meaning. This might very well depend on the particular aspect. So, let us take this example.

(Refer Slide Time: 09:02)

Do opinion phrases always have the same sentiment?

'Large' – positive or negative
Large screen vs. Large battery

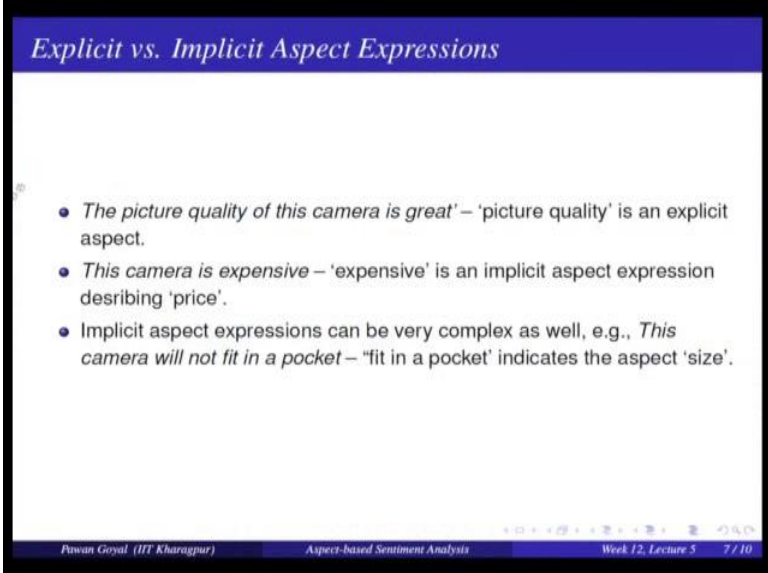
'Long' – positive or negative
Long battery life vs. Long loading time

Pawan Goyal (IIT Kharagpur) Aspect-based Sentiment Analysis Week 12, Lecture 5 6 / 10

So, I have this word large, large by itself we might think it is positive, but it can be positive or negative so these are two examples. So, take large screen versus large battery. Large screen you might like right my phone has a larger screen that might be positive sentiment, but large battery might be like it is becoming very hobby it will be a negative sentiment. Similarly with long it can be a long battery life it is always good right, but if you say long loading time it might not be a good; so same word might have different sentiment with different aspects.

You might have talk about opinions specific to the aspects then there is another problem sometimes the aspects are mentioned explicitly. So, once you identify some aspects you can then find out whichever document they occur in whichever sentence they occur in you take that already, but sometimes they are mentioned only indirectly like the (Refer Time: 09:59) we saw. Price is a direct aspect, but if you say it was expensive it is indirect. So, it is again an interesting problem how do I find out what are the indirect aspects, implicit aspects.

(Refer Slide Time: 10:11)



Explicit vs. Implicit Aspect Expressions

- The picture quality of this camera is great – 'picture quality' is an explicit aspect.
- This camera is expensive – 'expensive' is an implicit aspect expression describing 'price'.
- Implicit aspect expressions can be very complex as well, e.g., This camera will not fit in a pocket – "fit in a pocket" indicates the aspect 'size'.

Pawan Goyal (IIT Kharagpur) Aspect-based Sentiment Analysis Week 12, Lecture 3 7 / 10

So, like here the picture quality of this camera is great, picture quality is an explicit aspect, but if I say the camera is expensive; expensive is an implicit aspect that corresponds to price. And sometimes then you have to be very very complex also ok.

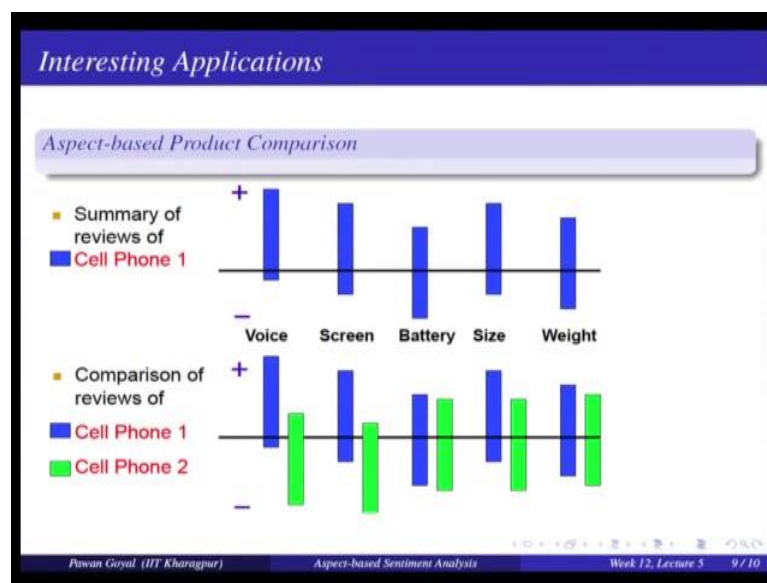
Like someone might say this camera does not fit in my pocket at all. So, saying will does not fill my pocket is saying it is very very this very huge size is huge. It is indicating the aspect size and that you are saying by some complex sentence.

So, identify these that there is some implicit aspect together is also an interesting research problem. Once we have this aspect base sentiment analysis you can now use it for many many different applications. So, what will one nice application for that? So, think about a website like Amazon or Flipkart where different people are talking giving reviews about a particular product, it is one thing to say overall the review a positive or negative, but suppose you want to go further you can say about this product what are the interesting aspects like camera, the price, and the image quality and so on.

In each price what are the overall sentiment that is one thing, but suppose you can also say- what are people talking about price. What are people's comments about price, what are the people comments about image? So, there you want to gather all the sentences they talk about a particular aspect and then do some sort of summarization on top of that. That is why the summarizing methods that we talk in this course will be helpful. So, you say about the picture that is what people are saying, about the image quality that was that is what people are seeing, about the about the size of the camera that is what people are saying. So, this field is called aspect based sentiment summarization, opinion summarization. We are taking opinions and trying to summarize it for each and individual aspect. So, it will be like that.

So, you are taking reviews you are seeing these are the aspects like touch screen and voice quality within each you say these are the positive exam reviews is the negative reviews and then you summarize the positive and negative reviews separately for each of the aspect. Then another interesting application would be aspect based product comparison.

(Refer Slide Time: 12:28)



So, what are the different product comparison do you see in on the websites. So, the websites you see they are comparing products based on their specs. This has this much resolution and this is the size and so on. But they are not comparing on many interesting aspects like what is the voice quality of this mobile ok. They will not talk about that in those terms. So, by using the reviews of 2 products I can find out the common aspects and then try to compare them on those common aspects. Comparison can be again this has got this much score also I can talk about some summing.

So, this product p 1 is better than product p 2 in terms of this aspect because that is what customers are saying. And this will be a very nice application that you are able to say why should you prefer to go for product 1 than product 2 because when I read the review about this aspect that is what people are commenting and you are able to compare the 2 products. So, something like this. You can give summary reviews of one product like cell phone or you can compare to cell phone 1 cell phone 2 and you can see in terms of voice this is more positive, this is more negative, screen this is more positive, this more negative, battery they are of the same size.

Cell phone 1 is more positive than cell phone 2. So, like that you can do now once you have taken the aspects you found out the aspects from the various reviews. And there are

many many other things that you can do about sentiment. So, it is a huge topic. So that finishes our discussion on sentiment analysis and opinion extraction. As I was saying, we have only touched upon this field giving you basic intuitions, what are the simple methods you can use to at least get it started. And that is what we have done from many most of the application that we have talked about.

We talked about classifications, summarization, anterior linking, and information extraction so we did not cover them in very large details we only give you a intuitions and how simple things that you learn in the course can be helpful in solving these problems. And all of this I guess you would have got some nice idea.

(Refer Slide Time: 14:45)

Conclusion

Many more NLP Applications

- Machine Translation
- Question Answering
- Cross-lingual Applications
- Text Processing for social media, informal text

Deep Learning Techniques are being applied for most of the tasks.

Pawan Goyal (IIT Kharagpur) Aspect-based Sentiment Analysis Week 12, Lecture 5 10 / 10

But there are some many other application that we have not covered. One particular application for example, is machine translation. So, you take sentence from one language and you are trying to translate it to another language. Again so different techniques that we have used can be applied plus there are some other techniques that we will be used like, how do you align words in two languages and so on. Then there is nice application of question and answering this is also becoming important. You have a lot of text in terms of Wikipedia and all and you guys asking a query and you do not just want

irrelevant document, but you want an actual answer from those. So, this is also very interesting application, but it is very very challenging right now.

Then you can talk about various cross lingual application right. This is again very crucial right now, because you are having your information in one language like say English, but you are typing in Hindi. You are typing in a different language. So, by typing in Hindi how do you get information that is there in English? What kind of cross lingual method that you can use, again they will use various translation methods or you can use many other insights like using dictionaries that translate from one language to another language and so on. Again this is a nice application and then with a lot of social media coming in coming up applications on how do you process the language of social media.

So, how do you handle all these informal text? Yes; so how do you handle that, then how do you handle all this code switching and code mixing. What do I mean by that, when you write on twitters suppose you are a bilingual you use English as well as Hindi you start writing in English, suddenly you switch to Hindi. And you might mix words of English in Hindi in the same sentence. These are called code mixing and code switching depending on certain criteria. If I say code switching is one way here you change one language to another language completely at certain point and code mixing is where you keep on mixing words. So, it not directly completely switches one language to another.

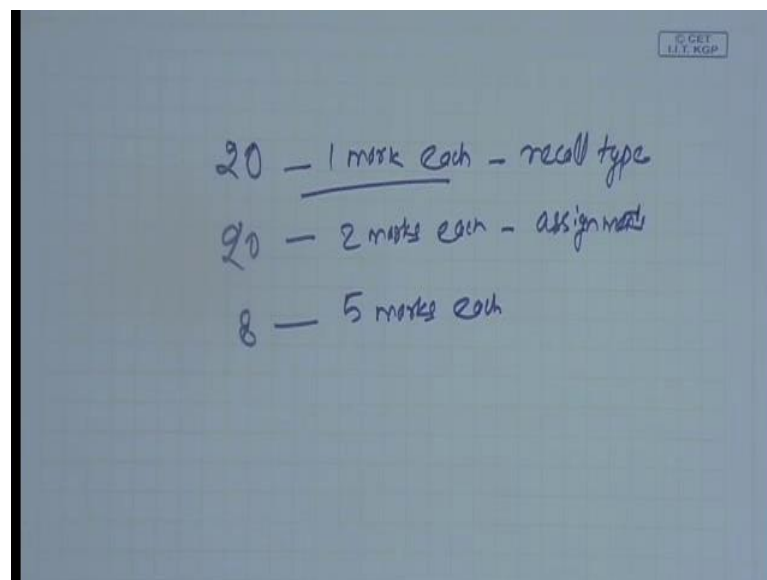
So, how do you process a language how do you find out what word belongs to what language how do you do all this part of speech tagging and say sentiment analysis over this kind of language is another interesting research direction that is coming up and there are many other application areas. Right now in many of this cases deep learning techniques are being applied, so we talk very briefly about this by word vectors, but many different deep learning methods are being applied, but that requires a complete different sort of course, where you are talk, you are thought about dual networks and this deep learning method.

But whatever techniques we have talked about in this course will be very very helpful when you would go and solve these sorts of problems and many of the application. And

even when you are taking any application when you are start reading different research papers you will again see that the basic the basic that I have covered in this course will be quiet helpful in understanding those research papers.

So, I think with that I will end this week and also this course. So, I hope that you found these 12 week in different course to be very very interesting, this you learnt a lot about this field of text minding and an N L P even the assignments probably helped you a lot in getting familiar with different tools and you feel confident, now that given a new problem that requires having some (Refer Time: 18:18) background you can try to tackle that. That was the main objective of this course and I hope that we have. So, you have been somewhat successful in that. So we will have an exam for this course, for that exam. Again the syllabus is the whole course that we have covered. All the slides will be there they want be any programming kind of questions in the exam. If I give you some idea, so this will be as per the format of NPTEL.

(Refer Slide Time: 18:52)



As per the format there will be first there will be 28 questions in the question paper. Dividing 20 20 8. So, 20 questions will be 1 mark each there will be like recall type questions. So, this will be coming from the slides. So, you should be able to recall certain

things that we have talked about in the slide. Then 20 questions will be 2 marks each they are based on the assignments.

So, they will be coming based on different questions that you solved in your assignments, again there will not be programming kind of questions and then there will be 8 questions of 5 marks each these are some questions that were probably not covered in the assignments and there will be some questions that we will require you to spend some more time, but hopefully you will be able to do that within that time that is allotted to you. So, one important thing that all these questions; most of these questions are like having only a single answer. So, they are multiple choice only a single answer is correct with 2 or 3 questions you might have to give the actual answer; so by in numerical as a numerical value.

So, this will happen with 2 or 3 or four questions. But if you have done this course, you should be able to solve to attempt all this problems and solve most of this. So, I will say, let me wish good luck for the further exam and thanks for being with us for the course and all my best wishes.

Thank you.