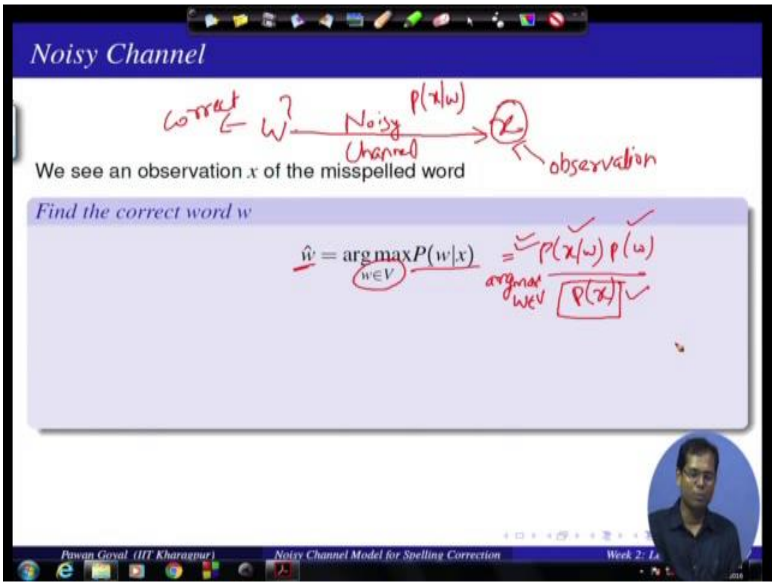
Natural Language Processing
Prof. Pawan Goyal
Department of Computer Science and Engineering
Indian Institute Technology, Kharagpur

Lecture - 08
Noisy Channel Model for Spelling Correction

Welcome back for the 3rd lecture of this week. So, in the last lecture, we talked about what are the various variations of edit distance that you might be using in general. So, today we will now talk about very extended algorithm for doing spelling correction.

So, in general when I talk about spelling correction, I might be talking about isolated word reduction or correction that is a word without the context, even there it might be a word that is in my vocabulary or not in my vocabulary. So, we will see examples for that or I might talk about doing correction in the context, there again a word might be in the vocabulary or out of the vocabulary and we will see how the simple approach of noisy channel model can be applied for doing a spelling correction.

(Refer Slide Time: 01:19)



Noisy Channel

$$\text{correct} \leftarrow w? \xrightarrow[\text{Channel}]{\text{Noisy}} \boxed{x} \xleftarrow{} \text{observation}$$

$P(x|w)$

We see an observation x of the misspelled word

Find the correct word w

$$\hat{w} = \underset{w \in V}{\arg\max} P(w|x) = \underset{w \in V}{\arg\max} \frac{P(x|w)\,P(w)}{\boxed{P(x)}}$$

Pawan Goyal (IIT Kharagpur)
Noisy Channel Model for Spelling Correction
Week 2: Le

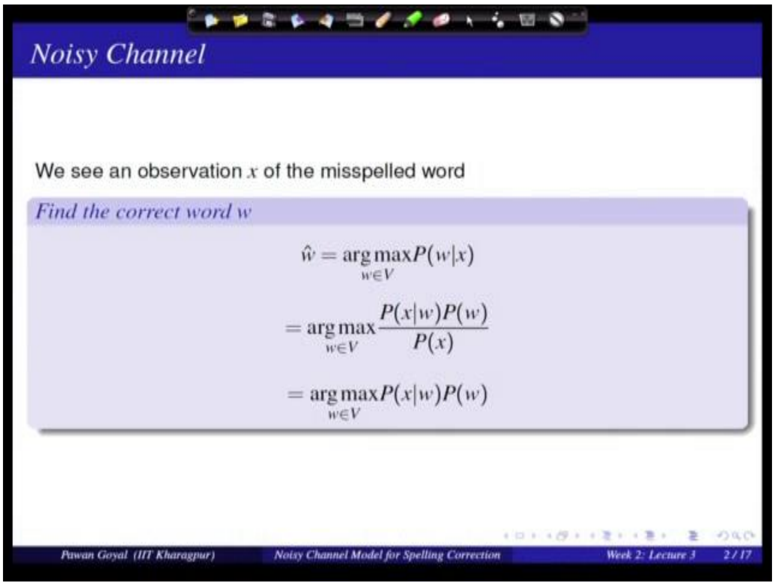
Now, what is my noisy channel model? So, if you think of the model name, what is happening? We are having a observation x that is a word, that is misspelled and you wanted to correct it that is your observation, but what is the idea of this model? So, whenever have a misspelling you wanted to write a correct word and you went through a channel that can be your keyboard or something else and that is how your word got misspelled. So, that is what you are assuming you have a correct word w, this is your correct word and you have a incorrect word x that is your observation, yes, now you wanted to type this correct word w, but you went through a channel and this we are saying is a noisy channel and you are making some mistake while going through this channel and the word that you are observing is x, but not w. Now your observation in x, now given this observation x how do I find out what is my mostly slightly candidate for w? What is my w?

In general how can I try to up to this problem? I am seeing a word x that I know it is misspelt, now how do I know it is miss spelt? 1 crude way of doing that is I maintain a dictionary that is my set of all the words my vocabulary and if the word x does not match with any of this word; I say this might be in misspelled. Now once I know x is a misspelled word; that means, I have to correct it, now what are I have to first start with what are all the possible candidates for correction. So, suppose possible there could be I find out all the words within the added distance of something now among those which one is the most slightly candidate. So, so in terms of noisy channel model what is the w from which x has been written. So, missspelt and we wrote x. So, in terms of noisy channel model I want to find out W hat that gives me the maximum probability P w given x among all the possible words in my vocabulary or it can be in my set of candidates what is the what that gives me the maximum value for P w given x.

Now, how do I compute this probability P x given x? You see as for the noisy channel model I first started with the word w and then I went through the channel and I ended up with the word x. So, I can only find out using my channel what is the probability of x given w, I will have to somehow used to find out probability of w given x yes. So, what is a particular theorem in probability that you can use for is. So, remember Bayes theorem. So, I want to find out probably w given x, but I already can find out from my channel probability of x given w. So, how do I write p w x in this term? So, I say this will be P x given w, P w divide by P x yes. So, I can replace I by arg max of this.

Now what is your variable here? The variable is w you can have multiple different words, but your x remains the same for each individual word. So, what, in fact, you are trying to do? For each word you are completing this value and then you are seeing that is the maximum value which word gives you the maximum value. Now among the 3 different probabilities that you are using for this formula which one you might remove for this computation? Is it the probability P x, remains the constant across all the words. So, it will not matter to my decision of which one is having the highest probability because this was the simply depend on this particular multiplication. So, I can as well remove P x and keep my argument and that is how we actually use.

(Refer Slide Time: 06:03)



Noisy Channel

We see an observation x of the misspelled word

Find the correct word w

$$\hat{w} = \arg\max_{w \in V} P(w|x)$$

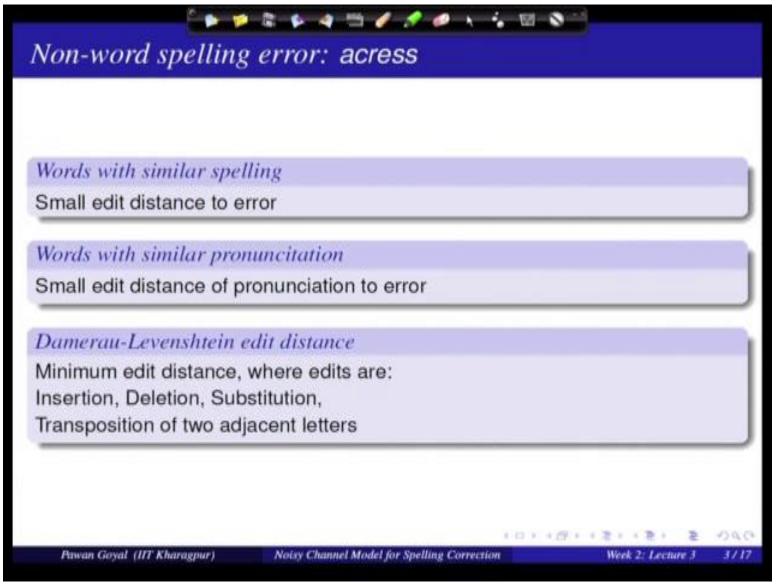
$$= \arg\max_{w \in V} \frac{P(x|w)P(w)}{P(x)}$$

$$= \arg\max_{w \in V} P(x|w)P(w)$$

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 2 / 17

I write it like that and I remove the probability P x and that is how that is. So, how I will find out which one is the correct word w for x which give me the maximum probability for probability x given w times probability w. So, now, we will see how do we compute these probabilities individually what are the various which we can compute this probabilities.

(Refer Slide Time: 06:30)



Non-word spelling error: acress

Words with similar spelling

Small edit distance to error

Words with similar pronunciation

Small edit distance of pronunciation to error

Damerau-Levenshtein edit distance

Minimum edit distance, where edits are:
Insertion, Deletion, Substitution,
Transposition of two adjacent letters

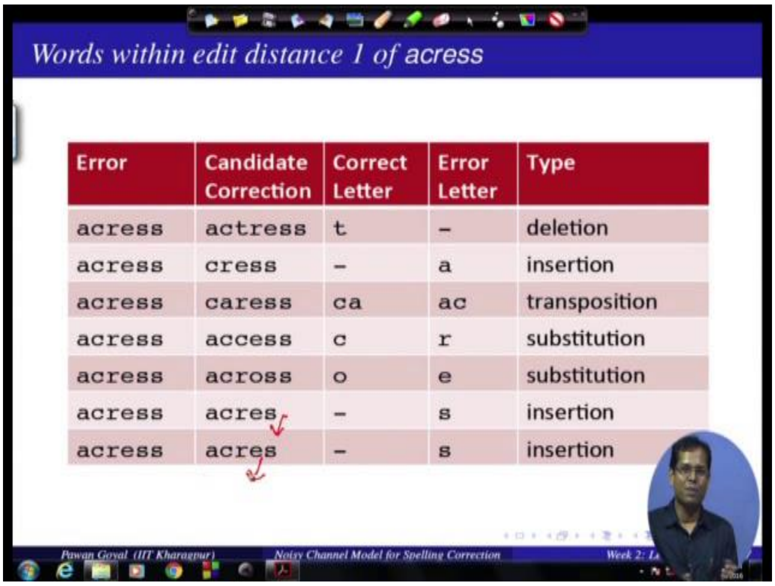
Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 3 / 17

So, we are taking about non word spelling errors. So what do I mean by non word spelling errors? A word that is not in the dictionary or my vocabulary and I know this might by misspelled word, suppose I have taken the word here actress. So, actress is not in the dictionary it can be actress across or whatever, but we do not know. So, this is the observation that we made, now using noisy channel model I want to find out what should have been the correct word. So, now, the first thing I will needed to do I will need to find out what are the candidate words that might be correct.

How do I find the candidate words I will try to find out words that are having similar spellings that are having small edit distances or I might try to use the words that are having similar pronunciation? So, both of these are possible. So, then I make some sort of candidate set. So, suppose I have ten different words that might have given to actress. So, these are my various candidates for W and I need to use the noisy channel model to find out which one is the most slightly candidate.

And suppose I am using might d l at a distance this is small variation of my Levenshtein distance remember Levenshtein distance what were the operations, we at insertion deletion and substitution, in this particular edit distance we also have the transposition. So, insertion, deletion, substitution and transposition of 2 adjacent correct words. So, we discussed about how to how to modify our dynamic problem algorithm to take into count of this transposition in the last lecture. So, now, so given 2 strings, I can always find out what is the word. So, given a string I can find out what are the other words that come within additions of 1 or 2.

(Refer Slide Time: 08:36)



Words within edit distance 1 of acress

Error	Candidate Correction	Correct Letter	Error Letter	Type
acress	actress	t	-	deletion
acress	cress	-	a	insertion
acress	caress	ca	ac	transposition
acress	access	c	r	substitution
acress	across	o	e	substitution
acress	acres	-	s	insertion
acress	acres	-	s	insertion

Pawan Goyal (IIT Kharagpur)
Noisy Channel Model for Spelling Correction
Week 2: L

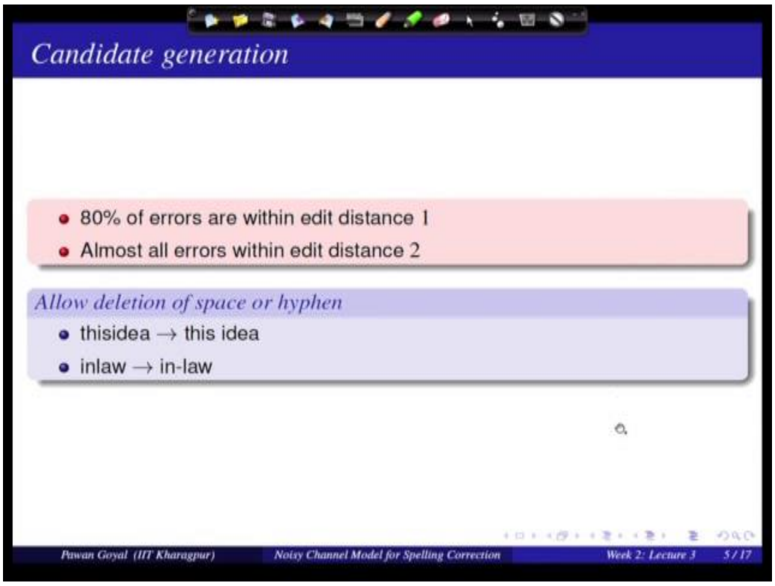
Suppose I start with acress that is my observation; incorrect word, I try to find out that are the also candidates are come with additions of 1.

Here is I have, here is the word acress, now I find the words like actress, cress, caress, access, across, acres, and acres occurs twice in this list, why and we will see that. Let us try to understand first candidate word actress. So, what does we have written here? So, we have written. So, this is my correct word this is my incorrect word. So, what change did I make for going from correct word to incorrect word? So, I went I had a letter t in my correct word, but I replace it with null. So, this is a deletion, I delete it; t from my correct word. So, that is why this is deletion.

Similarly, how would I go from cress to actress, I have to have insert a. So, this is the again insertion, yes, similarly, how I go from caress to acress I am doing a transposition. So, ca change to ac transposition - access to acress; c got change into r across to acress o got change to e now. Why there were 2 acres? you see acress I can go by doing 1 insertion, but insertion can happen at 2 points it can happen either here or it can happen here and these 2 will have different probabilities. remember with the probabilities of substitution will depend on what is the pervious word that what is the pervious character in this word. So, they will different probabilities, that is why they are 2 different insertions.

Now, that means, whenever I have a misspelled word I have to first generate all the possible candidates. So, there we try to use some sort of a statistics how many different candidates should I be using. So, the general statistics is that 80 percent of the errors are within distance of one and nearly all the errors are within the additions of 2 so; that means, I can mostly try to use the candidates within a additions of 1 or 2 may be, one will do for most of the cases.

(Refer Slide Time: 11:24)



Candidate generation

- 80% of errors are within edit distance 1
- Almost all errors within edit distance 2

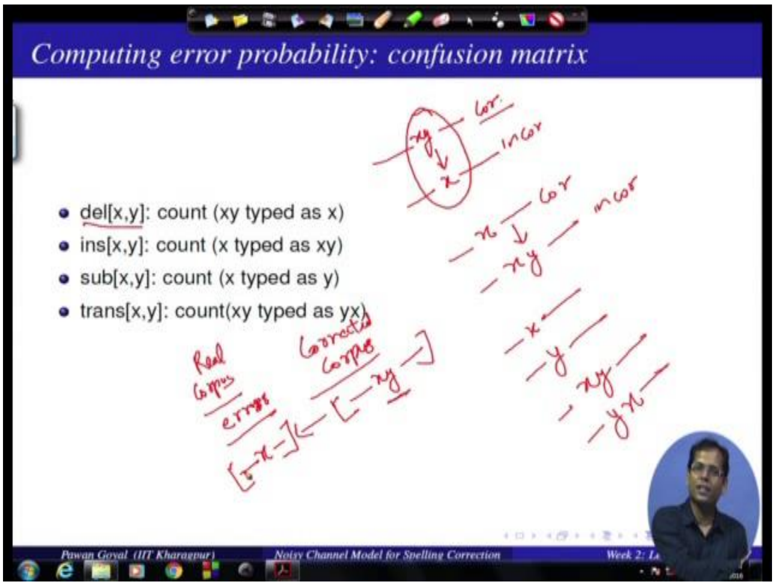
Allow deletion of space or hyphen

- thisidea → this idea
- inlaw → in-law

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 5 / 17

That is what we did in the previous example starting from acress we found out all the candidates with in additions of one and while generating the candidates I might also allow for deletion of space or hyphen. So, if I a word like this idea without a space, I might allow this idea with this space and similarly for hyphens and in law to in-law with a hyphen.

(Refer Slide Time: 11:45)



Computing error probability: confusion matrix

- del[x,y]: count (xy typed as x)
- ins[x,y]: count (x typed as xy)
- sub[x,y]: count (x typed as y)
- trans[x,y]: count(xy typed as yx)

Now when we are trying to compute the probabilities, we had 2 different terms probability of x given w and probability of w. So, you will first see how do we compute probability of x given w. Now suppose we are taking the simple case where there are only one this only one edit operation from the input string or the correct word to the incorrect word.

What do I mean by probability of x given w? I would mean what is the probability of that edit operation being done. So, remember we had four operations, insertion, deletion, substitution and transposition. So, we need to define probabilities for each of these. So, here that is what we are defining we are defining deletion over x y what do I mean by that? Whenever in the input string a input word I had the sequence x y this is my correct and it got typed as x this is my incorrect. So, x y got typed as x. So, this is I have deleted y. So, this will depend on the pervious character x in the correction.

Similarly, what about insertion I had x in the correct and I inserted y in the incorrect. So, this is the case of insertion substitution x got typed as y. So, I had x in my correction I got typed as y in my in correct decision. What is transposition? Again x y here incorrect got typed as y x in my incorrect. So, these are the four different possibilities if I am having only one edit operation between the 2 strings. So, now, the question is how do I find out probability of a particular deletion or insertion operation. So, let us take one of these and see how we will actually find out find out these probabilities.

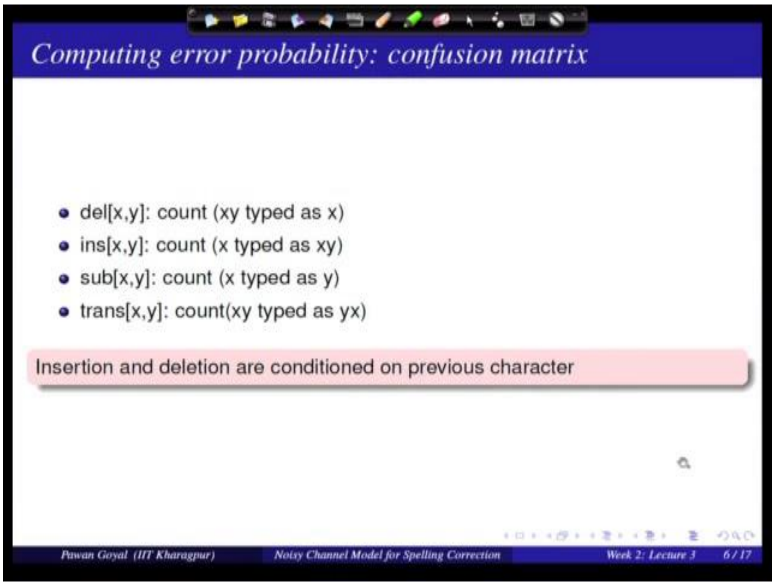
Let us take the simple case of deletion. So, how do we find the probability that in a string, a set of characters x y, a character bigram, I can also say it is a hecta bigram got converted to only x.

How I will feel find out this probability. So, we will have to see in general in corpus how likely is this error? So, now, what kind of data do I need to count to find this probability? So, I need some sort of data were you just have written some texts and they made some mistakes not knowingly. So, they made some mistakes that if general people do and somebody has then seen this corpus and tried to correct it. So, suppose I have a corpus that is corpus is a real corpus and that contains various errors and then I also have a corrected corpus.

Now, how will I use these 2 different corpora to find out the probability of deletion x y, I will say how many times do I seen my corrected corpus a word containing the bigram x y for which the corresponding word in my actual real corpus had only x; that means, actually the word should have been have having x y, but it was remains x. So, how will compute the probability? I will see in my real corpus how many times I have this corrected bigram x y and out of those how many times y was deleted in my real corpus and that will give me the probability that how after x what is the probability that y is deleted. So, whenever how many times x y occurred together out those how many cases y was deleted. And similarly I will find out find it out for all cases of insertion, substitution and transposition.

Let us see the actual matrices.

(Refer Slide Time: 16:23)



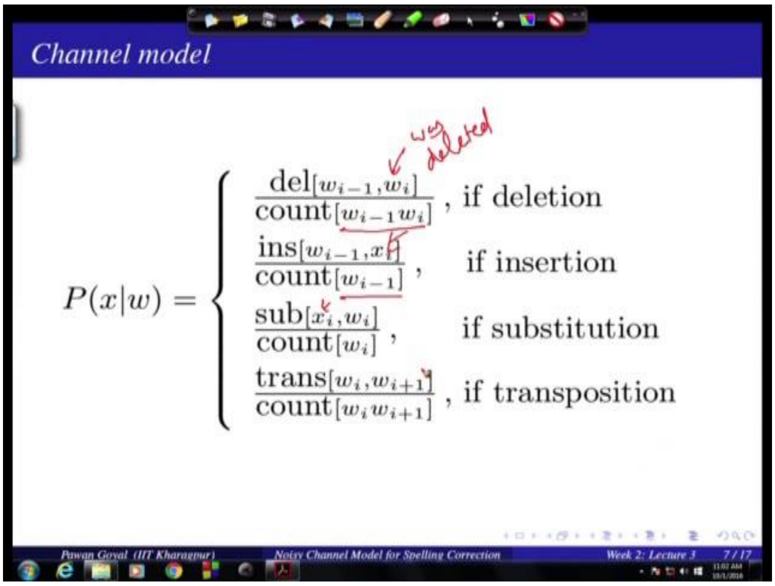
Computing error probability: confusion matrix

- del[x,y]: count (xy typed as x)
- ins[x,y]: count (x typed as xy)
- sub[x,y]: count (x typed as y)
- trans[x,y]: count(xy typed as yx)

Insertion and deletion are conditioned on previous character

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 6 / 17

(Refer Slide Time: 16:27)



Channel model

w_i deleted

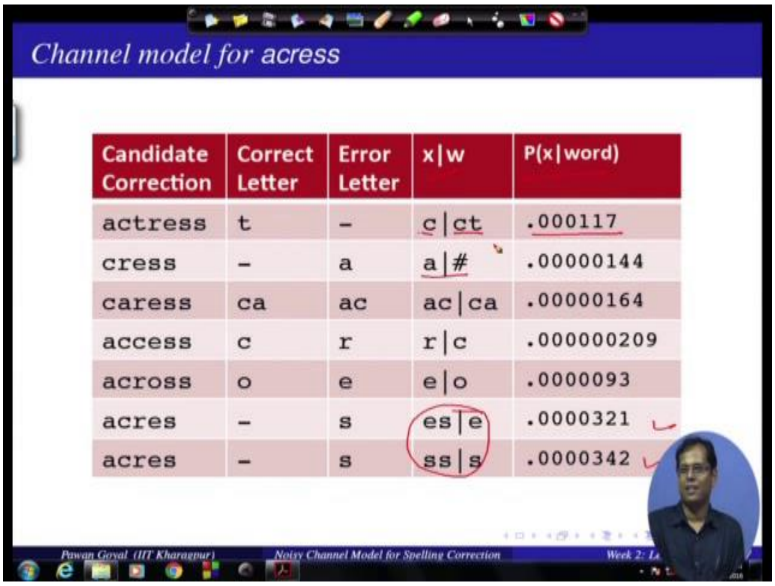
$$P(x|w) = \begin{cases} \frac{\text{del}[w_{i-1}, w_i]}{\text{count}[w_{i-1}w_i]}, & \text{if deletion} \\ \frac{\text{ins}[w_{i-1}, x_i]}{\text{count}[w_{i-1}]}, & \text{if insertion} \\ \frac{\text{sub}[x_i, w_i]}{\text{count}[w_i]}, & \text{if substitution} \\ \frac{\text{trans}[w_i, w_{i+1}]}{\text{count}[w_i w_{i+1}]}, & \text{if transposition} \end{cases}$$

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 7 / 17

So, you are seeing here insertion, deletions or conditioned of the pervious character. So, that is how I find out the probability of x given w x is my incorrect word and w is my possible candidate for correction. So, remember how we found out deletion, we said how many times this correct a bigram w i minus 1 and w i occur in my corrected corpus and among those how many times w i was deleted after W i minus 1. Similarly here insertion how many times the word w i minus one occurs in my corpus and among those how many times I inserted x i in my real corpus were people made mistake how many times they actually inserted x i after w I, w i minus q.

Similar substitution how many times w i occurs in my corpus, among those how many times it was substituted with x i. Same thing you will do for transposition how many times these 2 characters occurred together among those what fraction of times they were transposed. That means, we need 2 corpus - one we have made mistakes another the same corpus that has been corrected and then you can compute all these probabilities.

(Refer Slide Time: 17:58)



Channel model for acress

Candidate Correction	Correct Letter	Error Letter	x|w	P(x|word)
actress	t	-	c|ct	.000117
cress	-	a	a|#	.00000144
caress	ca	ac	ac|ca	.00000164
access	c	r	r|c	.000000209
across	o	e	e|o	.0000093
acres	-	s	es|e	.0000321
acres	-	s	ss|s	.0000342

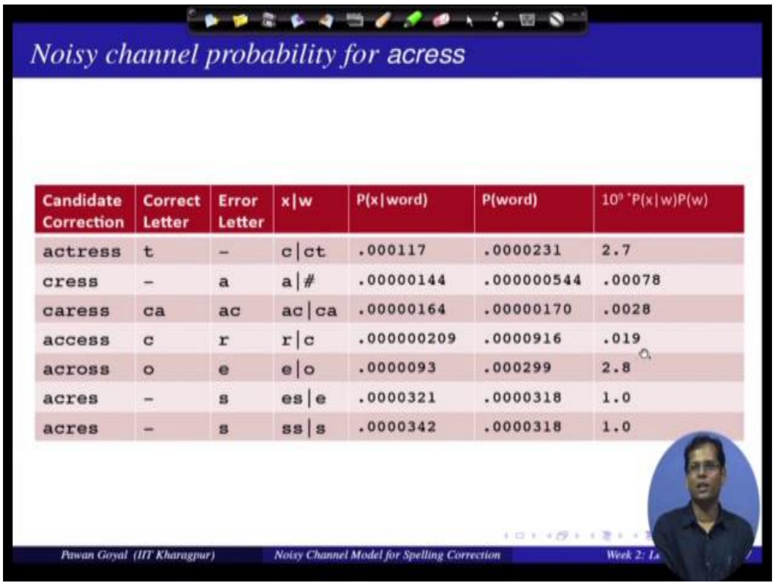
Pawan Goyal (IIT Kharagpur)
Noisy Channel Model for Spelling Correction
Week 2: L

Now, if we go back to the pervious example of finding various candidates for actress. So, we had 7 different candidates here.

Now, we already saw what is the corrected error word and we are using the model, we saw in the previous slide to find out probability of x given word this one. So, now, first we are writing what is my x given w? So, whenever this first my deletion from actress t got delete and we got actress so, how did you find the probability of deletion? We said what is the probability that ct, among all the times that ct occurs what is the probability that t gets deleted. So, out of ct, I see only c in my corpus and this probability, I find my corpus to be 0.000117 similarly insertion. That means, a got inserted in the beginning of the string transposition ac in place of ca and so on in so, here we have you see there are 2 different ways you can do, insertion after e or after s, that is why you have 2 different probability that is also.

Again from the corpus, you will find all these probabilities values; probability of x given my word for different words what is probability x given now what is the other probability of the compute probability of w itself. So, what is the intuition as such? You will try to give try to favor a word that is more likely to occur in the corpus then a word that is not so likely in the corpus.

(Refer Slide Time: 19:59)



Noisy channel probability for acress

Candidate Correction	Correct Letter	Error Letter	x|w	P(x|word)	P(word)	10^9 *P(x|w)P(w)
actress	t	-	c|ct	.000117	.0000231	2.7
cress	-	a	a|#	.00000144	.000000544	.00078
caress	ca	ac	ac|ca	.00000164	.00000170	.0028
access	c	r	r|c	.000000209	.0000916	.019
across	o	e	e|o	.0000093	.000299	2.8
acres	-	s	es|e	.0000321	.0000318	1.0
acres	-	s	ss|s	.0000342	.0000318	1.0

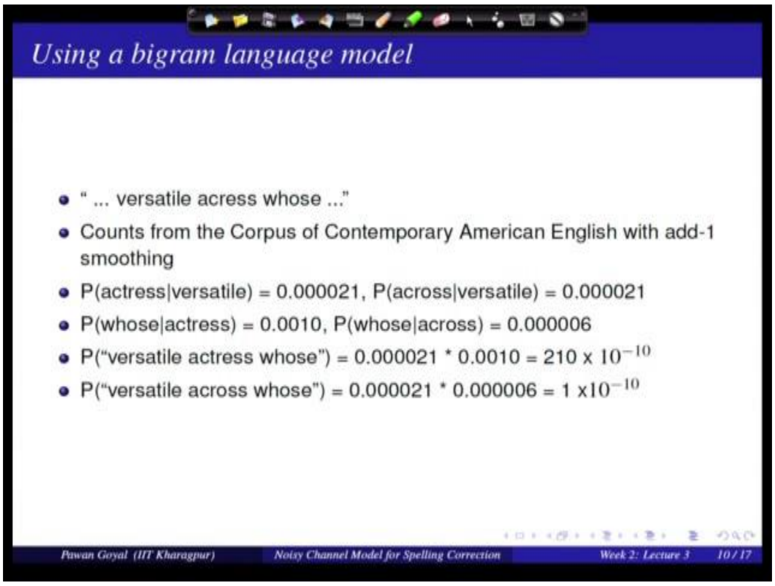
Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: L

Suppose I use a corpus and I also find out the probability of the word. So, now, I have found out probability of x given word and the probability of the word both of this. So, remember my noisy channel model I have to multiply these probabilities and find out which of the candidates gives me the highest probability value.

Suppose I do the same. So, in this slide you will see. So, I am multiplying these 2 probabilities and we are putting 10 to the power 9 just so that we are able to see these numbers they were be a 0.000 to the point that we will not able to compare. So, the first probability is 2.7 times 10 to the power 9 minus 9 sorry. So, what are the 2 candidates that have the highest probability? You see here, I have actress and I have across. So, these are 2 words that I having the highest probability and in general you might right across has the correct candidate.

Now remember we were saying, we can do it without with the context or with the context. So, without using the context this is the best we can do and suppose they come up to be really close 2.7; 2.71 without using the context you cannot find out which is the better candidate than the other, but suppose you are allowed to use the contexts. So, can you do better? So, let us see for the same spelling correction if I have the context what will I do.

(Refer Slide Time: 21:30)



Using a bigram language model

- " ... versatile acress whose ..."
- Counts from the Corpus of Contemporary American English with add-1 smoothing
- P(actress|versatile) = 0.000021, P(across|versatile) = 0.000021
- P(whose|actress) = 0.0010, P(whose|across) = 0.000006
- P("versatile actress whose") = 0.000021 * 0.0010 = 210 x 10^{-10}
- P("versatile across whose") = 0.000021 * 0.000006 = 1 x10^{-10}

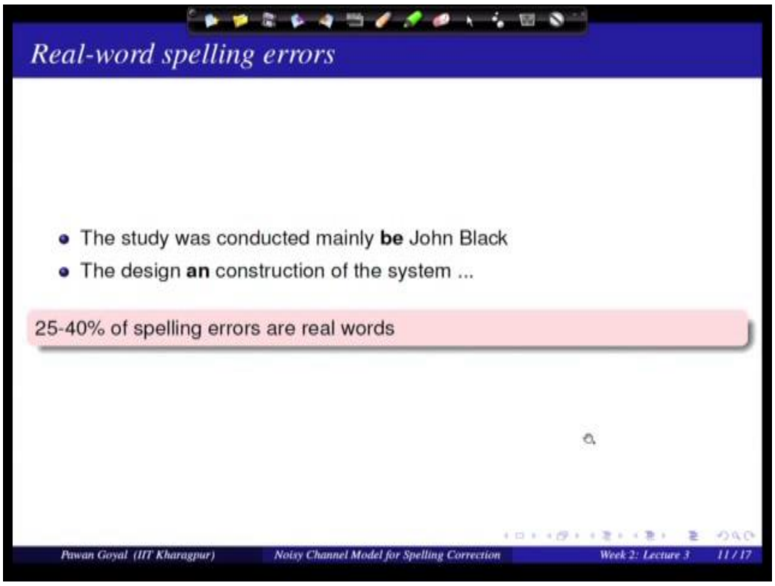
Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 10 / 17

Suppose my sentence says versatile acress whose and I have to correct it and remember what were 2 candidates? The 2 candidates are actress and across. So, the 2 sentences are versatile across actress, whose versatile across who is now which one is more likely.

Now, how do I find out? How do I use the fact that probably versatile actress or actress whose is more common than versatile across and across whose so. So for that what we can do we take some corpus. So, this is some corpus of contemporary American English and we use some smoothing. So, do not worry about what I mean by smoothing. We will cover that in detail in language modeling. So, suppose I do some smoothing, so idea is that if were we try to find out the probabilities of a word coming after another word that is what we are doing and from there I find out these probabilities that the that the word actress occurs after versatile is 0.000021 and the probability that the word across occurs after versatile is the same and that makes sense also, versatile across many different fields and versatile actress they might have the same sort of probability of occurrence in the corpus.

Now, what about the other bigram? What about the other 2 word combination that is actress whose and across whose? So, we see in the corpus we find actress whose has a much higher probability than across whose and if you have this information we can use that to refine the probabilities for that. If suppose I simply multiply the probabilities with my earlier corpus. So, what will I get? So, I will say probability of versatile actress whose is the initial the 2 probabilities 210 times 10 to the minus 10 and probability of versatile across whose will be 1 times 10 to the minus 10. So, this gives me that versatile actress whose is much more lightly than versatile actress, whose as per my corpus if I am using the context and this can further help me to find out what is the correct word that should have been there in place of actress.

(Refer Slide Time: 24:01)



Real-word spelling errors

- The study was conducted mainly **be** John Black
- The design **an** construction of the system ...

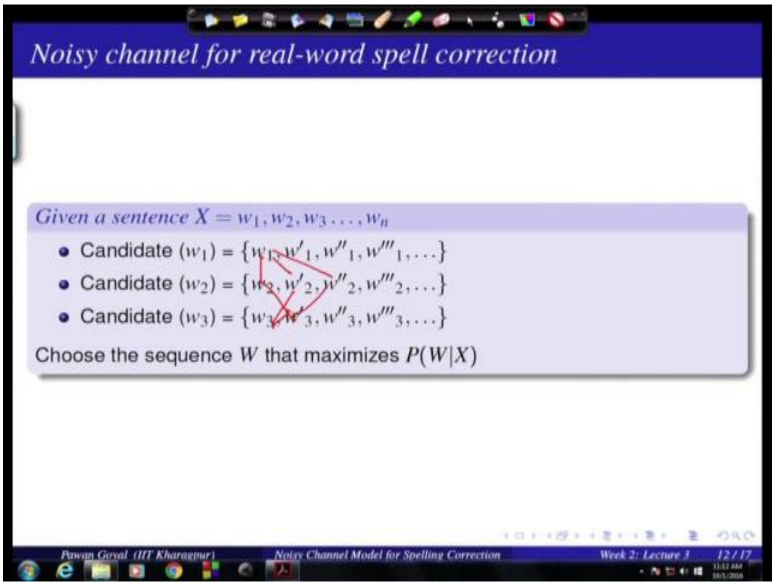
25-40% of spelling errors are real words

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 11 / 17

Now the next concept is the real word spelling errors. So, what do I mean by that? There might be errors where the misspelled word is actually in my dictionary. So, like you see the sentence here, the study was conducted mainly be John Black and we know the correct word should have been by, but the user ended up typing be. Now just think for a moment can I solve this problem without using the context. So, without using the context, how can I even say that it should be by and not be I might use the individual probability of a word occurrence, but that is not a good method to say that, so in that case, I might find errors everywhere in my corpus.

So, I should be try to using my context in somewhere. Same in the next sentence the design in construction of the system and I know this should be and, but the user ended up typing an now these are real words in the vocabulary and now I want to correct the spellings one thing is sure I should be using the context now how do I actually use the context for trying to correct these errors and this is again some statistics that 25 to 40 percent of the spelling errors that this here are real words.

(Refer Slide Time: 25:28)



Noisy channel for real-word spell correction

Given a sentence $X = w_1, w_2, w_3 \ldots, w_n$

- Candidate (w_1) = $\{w_1, w'_1, w''_1, w'''_1, \ldots\}$
- Candidate (w_2) = $\{w_2, w'_2, w''_2, w'''_2, \ldots\}$
- Candidate (w_3) = $\{w_3, w'_3, w''_3, w'''_3, \ldots\}$

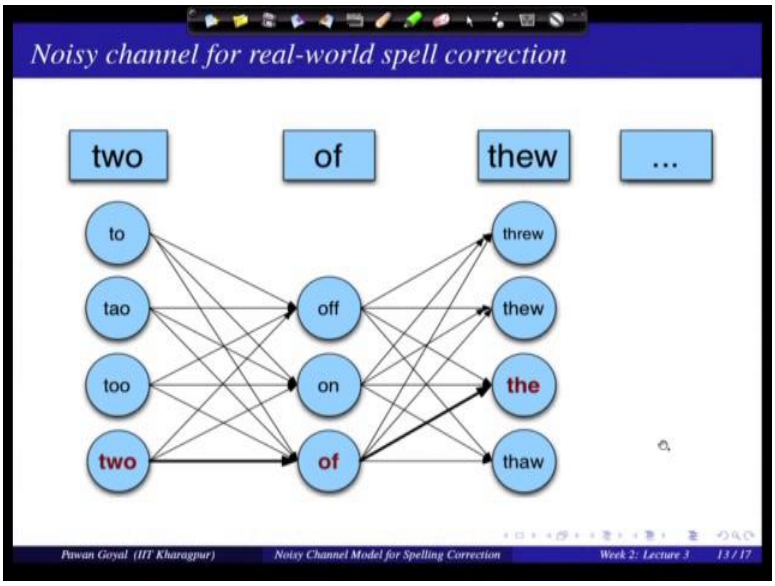
Choose the sequence W that maximizes $P(W|X)$

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 12 / 17

Now suppose I have a sentence X that contains words W 1 to W 1 now we do not know which of these words is in error because all of these are real words in my dictionary. So, how do I actually start doing the spell correction? So, what I will do for each of the individual word I will try to find out what are the best possible candidates that might be the correct words and there I should not forget to include this word itself because this might be a correct word also. So, here what we are doing for the word W 1, we are saying candidates are W 1 or some word W 1 prime W 1 double prime W 1 triple prime and so on, they are possible candidates for correction. W 2 again that is what W 2 and other candidates same way W 3, now what is my problem? Now I can have any sequence from the candidates of W 1, I can choose any one W 2, I can choose any one and so on and among all these possible sequences I have to find out which one is giving me the highest probability in somethings.

I want to find the sequence W that maximizes the probability of W given X that is whether this would be a good sequence or this would be a good sequence or this would be a good sequence or so on, which of this sequence will be the best in this case. So, again you have to find out W that maximizes probability of W given X. So, again can we try to use the noisy channel model? So, we will see.

(Refer Slide Time: 27:15)



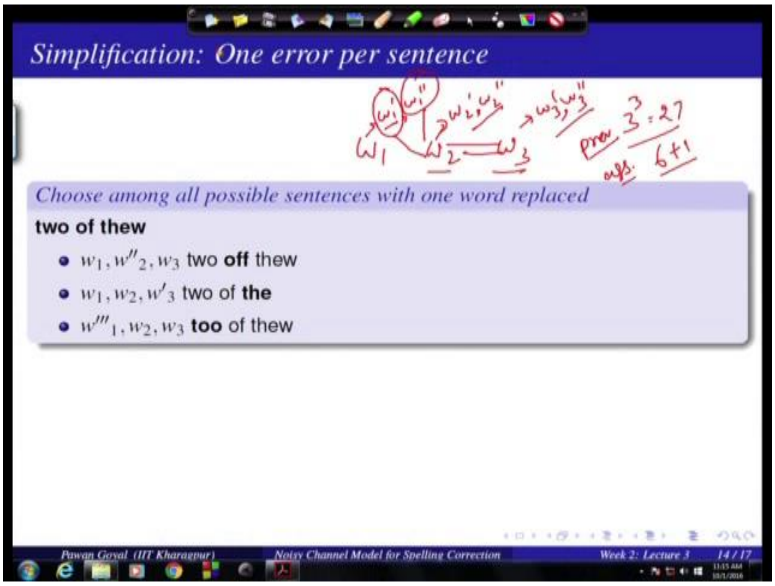
Noisy channel for real-world spell correction

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 13 / 17

Now we are taking a real example here 2 of w, thew and probably it should have been 2 of the; and we are not seeing the other words. So, as my; the model what will I do I will take the word two, I will put this word along with all the other possible candidates.

I have put in put t W o, t a o, t u, all are having a smaller distance similarly with off I am putting o n, o double f with the thew, I put thew, threw, thaw, the and so on and we know the correct sequences 2 off the fine, but in general suppose I had to find out the probability for each sequence this might become exponential. So, you can do the simple math suppose each word on an average has four possible candidates including itself and there are five different words in my sentence. So, how many different sequences I will have? 4 to the power 5 and if the number of words increases in my sentence this will keep on increasing.

(Refer Slide Time: 28:31)



Simplification: One error per sentence

Choose among all possible sentences with one word replaced

two of thew

- w_1, w''_2, w_3 two **off** thew
- w_1, w_2, w'_3 two of **the**
- w'''_1, w_2, w_3 **too** of thew

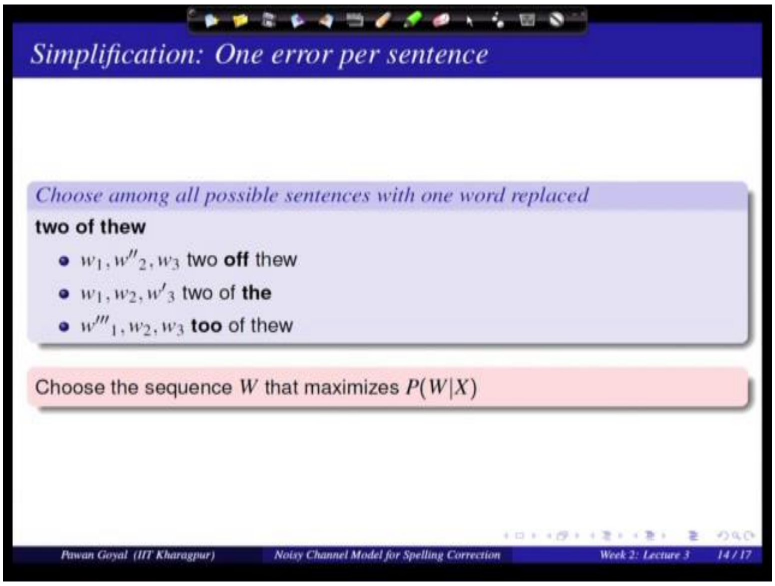
Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 14 / 17

Now to avoid this problem, we will try to use some sort of we make some assumptions that it is also simplification to the model what is that. We say let us assume there will be a at most one error in the sentence. So, how does it help? It helps in that whenever I am choosing the candidates, I will only choose a different word for one of the words. So, what do I mean by that? Suppose I had my sentence W 1, W 2, W 3 and this has some other possible things like W 2 prime W 2 double prime, this has W 1 double prime W 1 double prime and this has W 3 prime, W 3 double prime. So, in the previous model, we would have taken all the 3 cube that is 27 different candidates in the previous one.

So, with this assumption what we will do? For W 1, if I am choosing a candidate of W 1 prime, I will assume W 2 and W 3 were correct. So, this is the assumption there might be at most one error per sentence.

How many candidates will be there? I will have 2 possibilities here, 2 possibilities here and 2 possibilities here. So, there would be at most 6 plus 1, if you are taking the 1, there the original sentence might have been correct say W q W 2 W 3 as it is. So, the candidates are W 1 prime W 2 W 3 only one error per sentence W 1 double prime W 2 W 3 and so on. So, this hugely reduces the number of candidates for which I have to check and this is quite true also and that is why you do not make more than one error per sentence in general.

(Refer Slide Time: 30:18)



Simplification: One error per sentence

Choose among all possible sentences with one word replaced

two of thew

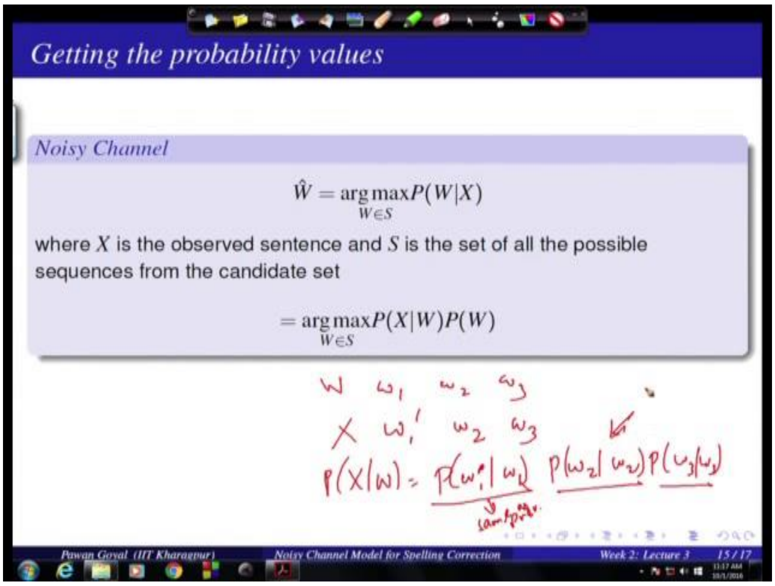
- w_1, w''_2, w_3 two **off** thew
- w_1, w_2, w'_3 two of **the**
- w'''_1, w_2, w_3 **too** of thew

Choose the sequence W that maximizes $P(W|X)$

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 14 / 17

Now once we have this assumption we want to find this sequence W that maximizes the probability of W given X. So, I can again try to use my noisy channel model.

(Refer Slide Time: 30:33)



Getting the probability values

Noisy Channel

$$\hat{W} = \underset{W \in S}{\arg\max} P(W|X)$$

where X is the observed sentence and S is the set of all the possible sequences from the candidate set

$$= \underset{W \in S}{\arg\max} P(X|W)P(W)$$

$W \quad w_1 \quad w_2 \quad w_3$

$X \quad w_1' \quad w_2 \quad w_3$

$$P(X|W) = \underbrace{P(w_1'|w_1)}_{\text{sam prob.}} \; P(w_2|w_2) P(w_3|w_3)$$

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 15 / 17

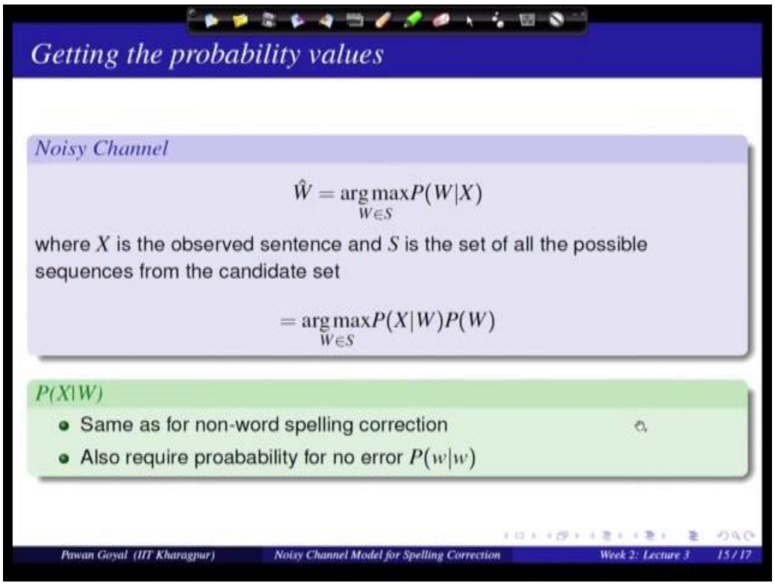
How find out the sequence W hat that maximizes the probability of P W given X, X is my observed sequence and W is one of my candidate sequences and we know how to find out the candidate sequences. So, that is for each of these sequence is W, I have to find out the probability W given X now as for the noisy channel model how will I write it down. So, I will again use the Bayes theorem here because I am starting from W and going to x. So, I can only find out the probability of X given W. So, this is what I will do now how do I write probability X given W X is a sequence W is a sequence. So, suppose they have the same number of words I can make a simplifying assumption that P X given W is simply multiplication of probability of individual word given that in the sequence.

So, suppose my word is this is my W and this is my x. So, probability X given W i will write X probability of W 1 prime given W 1 probability W 2 given W 2 probability W 3 given W 3 . So, now, there are 2 kind of probabilities here one is when the 2 words are not the same probability of doing an error yes one they at the same now how do I find out the individual probabilities this one we have already see in the previous case, I will find out what are the [audit/edit] edit operation and what is the corresponding probability if it is deletion insertion and so on.

This is same as the previous case same as in previous, but how do I find out the probability of W 2 given W 2 should it be 1, now it will depend on what is the kind of errors that you are seeing in your corpus if you are seeing that they are on an average one error per 10 words, you will say this probability is 0.9, thus this is what is correct these are probability of the same; the word is correct.

If there are 1 error in 100 words you will say that this probability is 0.99 and that is how you will fix this probability.

(Refer Slide Time: 33:02)



Getting the probability values

Noisy Channel

$$\hat{W} = \arg\max_{W \in S} P(W|X)$$

where X is the observed sentence and S is the set of all the possible sequences from the candidate set

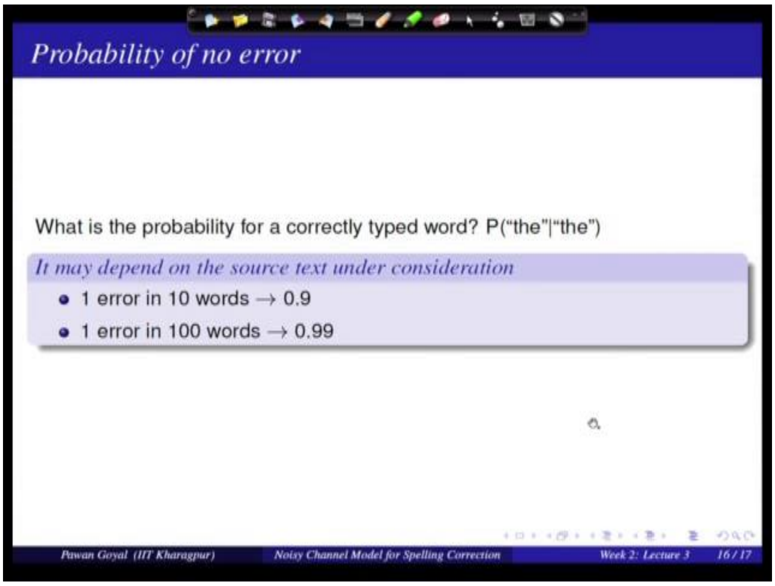
$$= \arg\max_{W \in S} P(X|W)P(W)$$

$P(X|W)$

- Same as for non-word spelling correction
- Also require proabability for no error $P(w|w)$

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 15 / 17

(Refer Slide Time: 33:11)



Probability of no error

What is the probability for a correctly typed word? P("the"|"the")

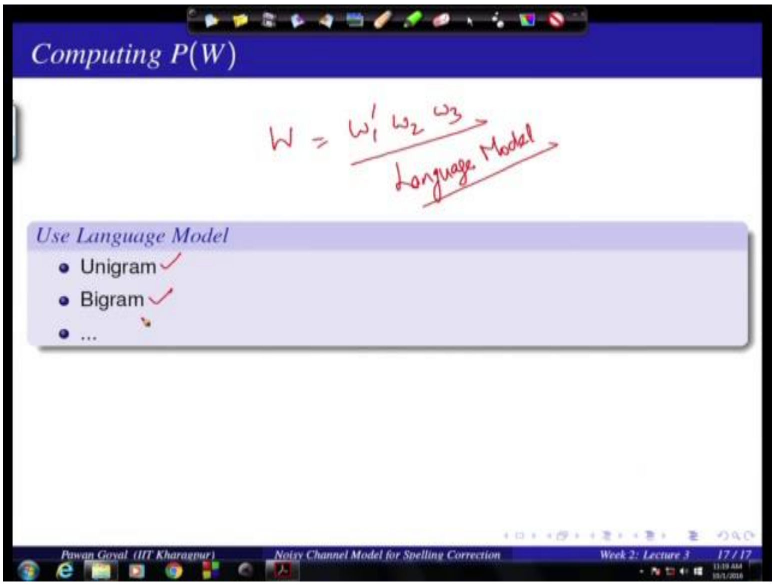
It may depend on the source text under consideration

- 1 error in 10 words → 0.9
- 1 error in 100 words → 0.99

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 16 / 17

P X given W same as non word spelling correction, but we also need this probability w given w and that is the probability that you have correctly typed a word and that you can find by the kind of the amount of errors you are seeing in your source. If there is one error per 10 words you will say this probability is 0.9, if there is one error per 100 words you will say this probability is 0.99 that is how you will set all these probabilities and you will choose the candidate X that gives you the maximum probability, candidate W sorry.

(Refer Slide Time: 33:37)



Computing P(W)

$W = w_1' \, w_2 \, w_3 \rightarrow$

$\xrightarrow{\text{Language Model}}$

Use Language Model

- Unigram ✓
- Bigram ✓
- ...

Pawan Goyal (IIT Kharagpur) Noisy Channel Model for Spelling Correction Week 2: Lecture 3 17 / 17

Now, the other part is finding out probability W by W I mean the sequence W 1 prime, W 2, W 3 how do I define the probability of the sequence W. For that we will use a technique called language modeling. So, that will give me the probability of n is the sequence of words and we will use any of this unigram model bigram model or any other model for finding out this probability of W.

This here will be the topic for the next lecture. So, in this lecture we discussed what are non word errors, re-word errors how do we use noisy channel model to correct those, but there we also need something like probability of the sentence as such probability of the sequence of words W, how do we obtain that and that is why we will go to the idea of language modeling in the next lecture.