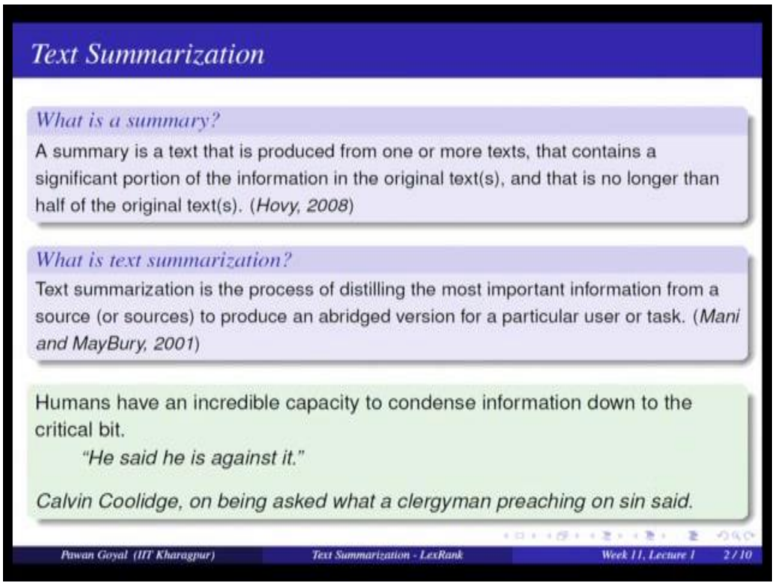
Natural Language Processing
Prof. Pawan Goyal
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 51
Text Summarization – LEXRANK

Hello, everyone welcome to the week 11 of this course. So, we were talking about various applications and this week we will focus on two different applications; text summarization and text classification. So, this lecture we will start talking about text classification and we will focus on one very important approach that is using LEXRANK algorithm for summarization.

(Refer Slide Time: 00:45)



Text Summarization

What is a summary?
A summary is a text that is produced from one or more texts, that contains a significant portion of the information in the original text(s), and that is no longer than half of the original text(s). (*Hovy, 2008*)

What is text summarization?
Text summarization is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user or task. (*Mani and MayBury, 2001*)

Humans have an incredible capacity to condense information down to the critical bit.
"He said he is against it."

Calvin Coolidge, on being asked what a clergyman preaching on sin said.

Pawan Goyal (IIT Kharagpur) Text Summarization - LexRank Week 11, Lecture 1 2 / 10

So, what is text summarization? So text summarization as the name would says that you have a lot of information with you and you want to produce a summarized form. So, you want to summarize this information to some reasonable extent. So, you are trying to distill information and you are trying to present only a summary out of that. So, there are some definitions that you can find for this task of text summarization.

For example, so then Hovy in 2008 gave this definition, so a summary is a text that is produced from one or more texts that contain significant portion of the information in the original text and is no longer than half the original text, so what do you see here?S

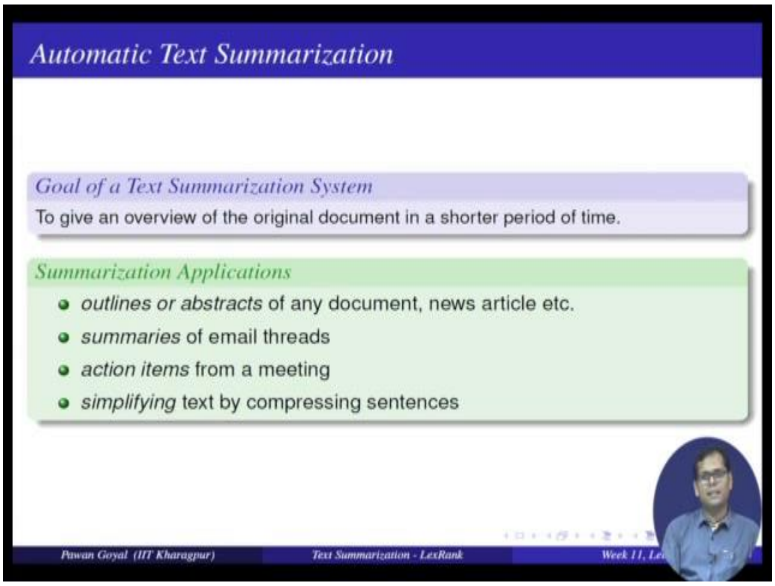
So, you have some initial data text data; it can be some documents one document set of documents or whatever and that contains some information, you want to convert that into some different form. So, that will be again text but now it is much more smaller than the original data. So, it can be may be half or even one-fourth of the original data; what is important is that, it should contain the most the important information from the original text. So, you want to retrieve the original important information from there and try to compress otherwise that is the goal of summarization.

So, we can also give this definition that it is the process of distilling the most important information from a source to produce an abridged version for a particular user or task. So, it might happen that you need a very generic summary or it might be for a particular user or a task, so that is also possible in the case of summarization.

So, now so when we talk about summarization, so you might have written summaries yourself when in your school days and all and it is said that humans are generally very good in doing summarization. So, they can not only reduce it to half, one-fourth they can also go down to the very critical bit what is the actual information that is there. So, this is like a; so Calvin Coolidge, so he heard a lecture for 4 hours on a clergyman preaching on sin and somebody then asked what did he say; so he can mention that in one bit that he; so that person said that he is against it. So, this is like condensing the information to the very critical bit, so humans are very very good at that.

But what about the machines, so what can machines do. So, what how can they achieve a summary from a given text.

(Refer Slide Time: 03:42)



Automatic Text Summarization

Goal of a Text Summarization System

To give an overview of the original document in a shorter period of time.

Summarization Applications

- *outlines or abstracts* of any document, news article etc.
- *summaries* of email threads
- *action items* from a meeting
- *simplifying* text by compressing sentences

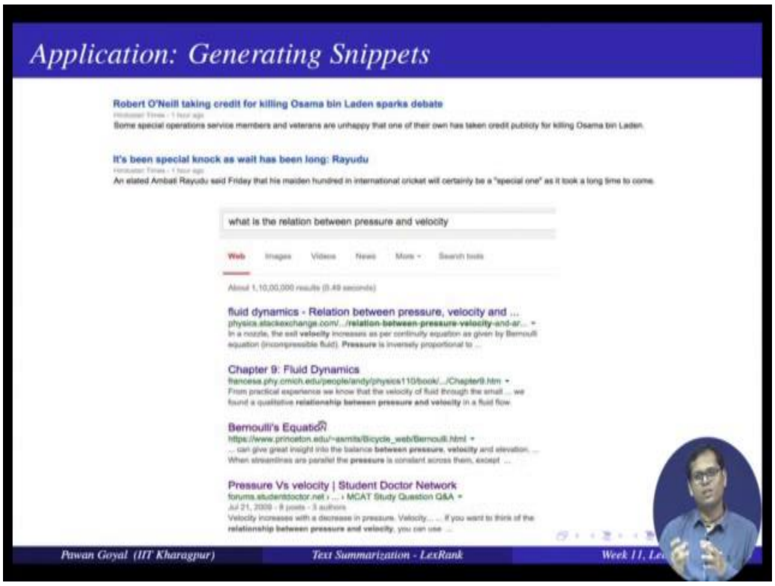
Pawan Goyal (IIT Kharagpur) Text Summarization - LexRank Week 11, Le

So, when we talk about automatic text summarization, so we are talking about an algorithm or a system that will take an input as a document or set of documents and for a task, it will produce a summary out of that. So, the goal would be to give an overview of the original document in a shorter period of time. So, we can see this definition with these different applications of summarization. So, for example, you are seeing suppose you want to read certain research article or a news article. So, it would be good if somebody can present a short summary of that user research article and then you can make your mind whether you want to read this whole article or not.

So, in scientific articles you get the abstracts as the summary, so abstracts tell us about what do you see in this research article, but news articles you have only headlines, you may not have the abstracts. Suppose you take news article and produce a summary of the what is the main information that is contained, that can also help you in saving time in if you want to read a lot of different articles or it can also help you understand whether you want to read it in more details.

Then you can talk about summaries of email threads, there are email threads between it may between customers and different agents for a particular product, it can be about a particular research task that you are having lot of conversations you want to find out what are the important points from this big email thread. We can talk about summarizing meeting minutes, section items for a meeting or you have a lot of sentences and you try to combine them and compress information from them. So, there can be many many different applications here that you can think of for text summarization.

(Refer Slide Time: 05:39)



Application: Generating Snippets

Robert O'Neill taking credit for killing Osama bin Laden sparks debate
Hindustan Times - 1 hour ago
Some special operations service members and veterans are unhappy that one of their own has taken credit publicly for killing Osama bin Laden.

It's been special knock as wait has been long: Rayudu
Hindustan Times - 1 hour ago
An elated Ambati Rayudu said Friday that his maiden hundred in international cricket will certainly be a "special one" as it took a long time to come.

what is the relation between pressure and velocity

About 1,10,00,000 results (0.49 seconds)

fluid dynamics - Relation between pressure, velocity and ...
physics.stackexchange.com/.../relation-between-pressure-velocity-and-ar...
In a nozzle, the exit velocity increases as per continuity equation as given by Bernoulli equation (incompressible fluid). Pressure is inversely proportional to ...

Chapter 9: Fluid Dynamics
francesa.phy.cmich.edu/people/andy/physics110/book/.../Chapter9.htm
From practical experience we know that the velocity of fluid through the small ... we found a qualitative relationship between pressure and velocity in a fluid flow.

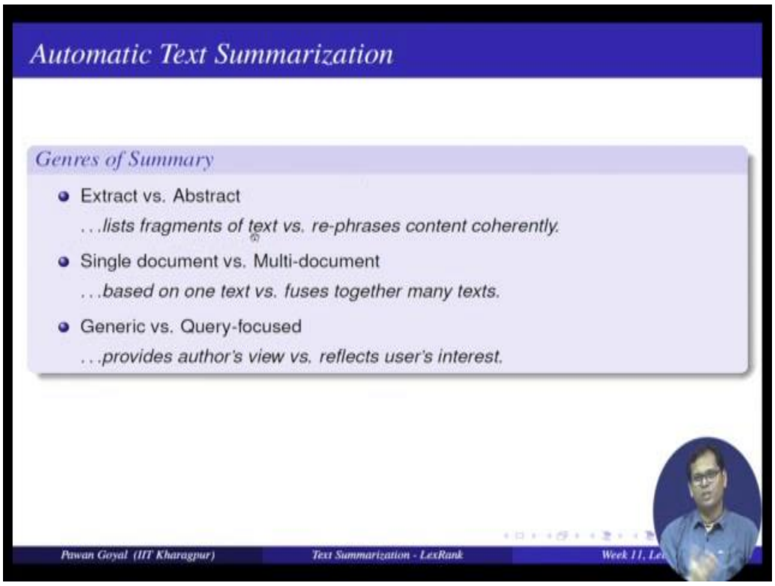
Bernoulli's Equation
https://www.princeton.edu/~asmits/Bicycle_web/Bernoulli.html
... can give great insight into the balance between pressure, velocity and elevation. ... When streamlines are parallel the pressure is constant across them, except ...

Pressure Vs velocity | Student Doctor Network
forums.studentdoctor.net › ... › MCAT Study Question Q&A
Jul 21, 2009 - 8 posts - 3 authors
Velocity increases with a decrease in pressure. Velocity... ... If you want to think of the relationship between pressure and velocity, you can use ...

Pawan Goyal (IIT Kharagpur) Text Summarization - LexRank Week 11, Le

Here is one application that you see like daily that is generating snippets from the web pages. So, you are searching for something you get some results, but with the results you also get some snippets; it is like here the news Robert Oniell taking credit for killing Osama Bin Laden sparks debate. So, this is from old news article and with that you also see a snippet from the news. Similarly if you do some search what is the relation between pressure and velocity, you get different results and with some results you can also get the snippet that is relevant to the question that you have asked and contains the important information from the document. So, this is another very important application for text summarization.

(Refer Slide Time: 06:25)



Automatic Text Summarization

Genres of Summary

- Extract vs. Abstract
  . . .lists fragments of text vs. re-phrases content coherently.
- Single document vs. Multi-document
  . . .based on one text vs. fuses together many texts.
- Generic vs. Query-focused
  . . .provides author's view vs. reflects user's interest.

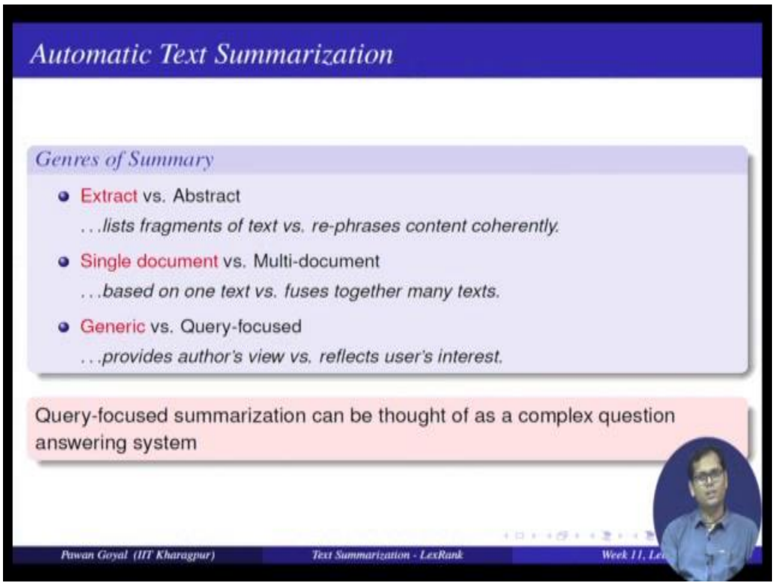
Now, when we talk about summary there are different genres in which you can produce summary. So, for example you can talk about building an extractive summary or an abstractive summary, so let us try to understand the difference.

In extractive summary, you will take the input document that and you will break it into say various sentences and you will try to pick up what are the important sentences from here, so you are extracting information from there. So, you are extracting some segments from there, but in abstractive summary you will also take information and rewrite in your own words and you will merge different sentences fuse them together and so on. So, in extractive summary you are listing fragments of text; on the other hand in abstractive summary you are taking content and rephrasing that so that becomes much more readable. So, when you are writing a summary you are probably doing it in abstractive manner.

Then you can have a single document summary or a multi document summary. So, you have a single news articles or you take multiple articles that are belonging to the same theme and you are producing the summary out of that, so both are possible. So, this is based on one text and this is fusing different text together, it can be generic summary from the input document. So, what is the generic view of the document or it can be a query focus summary that is you have got the query and you want to get a summary that is focused on your particular query.

So, the document might contain about contain many different informations but the information that is close to your query that will be summarized. So, you can also think it as a complex question answering task, you are asking a question the information is there in multiple documents, you find out information and reply to you based on what is relevant to your query.

(Refer Slide Time: 08:14)



Automatic Text Summarization

Genres of Summary

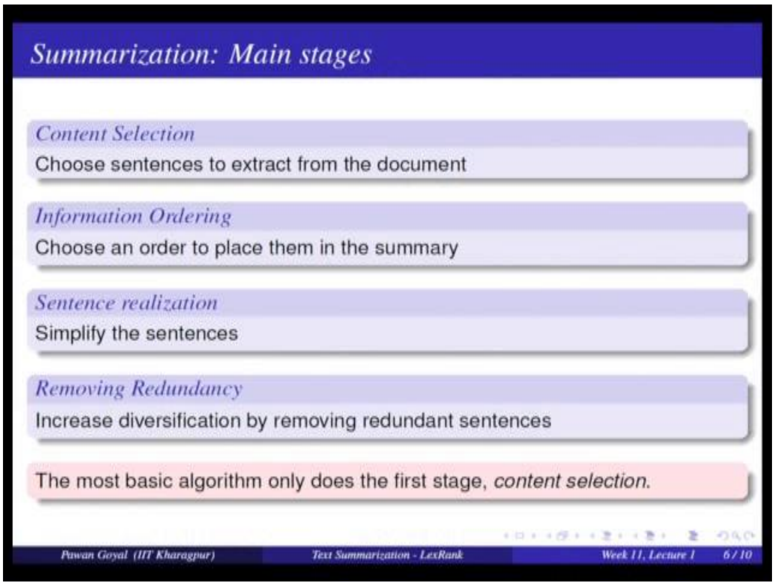
- Extract vs. Abstract
  ...lists fragments of text vs. re-phrases content coherently.
- Single document vs. Multi-document
  ...based on one text vs. fuses together many texts.
- Generic vs. Query-focused
  ...provides author's view vs. reflects user's interest.

Query-focused summarization can be thought of as a complex question answering system

Pawan Goyal (IIT Kharagpur) Text Summarization - LexRank Week 11, Le

So, in our next two three lectures we will mainly focus on extractive single document and generic summarization. So, we will talk about how do you take a document and create an extractive and generic summary out of that, but whatever method we will discuss are very easily extendible to the other forms like multi document summary and query focus summary and also for abstractive summary, we will give some discussion that how do you use this method for generating an abstractive summary. So, let us focus on the simplest task that can help us to talk about only the fundamentals that what are the different algorithms or hypothesis that you will use for constructing a summary from a given input document.

(Refer Slide Time: 09:06)



Summarization: Main stages

Content Selection
Choose sentences to extract from the document

Information Ordering
Choose an order to place them in the summary

Sentence realization
Simplify the sentences

Removing Redundancy
Increase diversification by removing redundant sentences

The most basic algorithm only does the first stage, *content selection*.

Pawan Goyal (IIT Kharagpur) Text Summarization - LexRank Week 11, Lecture 1 6 / 10

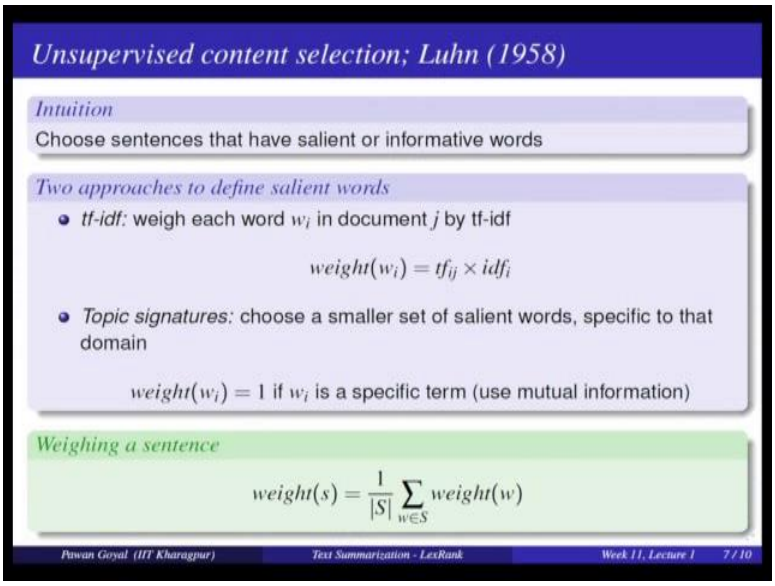
So when we talk about summarization what are the main stages that one need to go through so. Firstly, you are given an input document first task would be you will try to select the content that is important for summarization; that is find out what are the most informative sentences from here. So, if I am talking about extractive summarization; find out important sentences from this document. So, you will see how do we design an approach for finding the important sentences, what are the different hypothesis I can take.

Now, once I have selected the content; I know these sentences are important, what would be the next task? Next would be I will try to order them somehow. So, whether this sentence will go first or this sentence will go after this; this also depends on how will I make it more readable or more presentable to the end user; there I have to talk about how it can be more coherent and so on.

Then I might want to simplify the sentences that I am getting from the actual summary; that is I want to remove some of the redundant phrases and so on; this is like simplification. Once I have done that, I may also want to remove the redundancy that is I found what are the important sentences but there two of the sentences are very similar to each other. So, I want to remove a sentence that is similar to some of the sentence that is already present in my summary. So, I want do not want to give the same information multiple times because I have a shortage of space.

So, whenever I talk about summarization; I generally put a restriction that I want for a document I want a summary of length 100 or length 200. So, I do not want to waste the space by giving the same information multiple times, this is like getting also diversity in the information. So when we talk of the most basic algorithm, this will only concern the first part that is the content selection; how do I find the most important sentence in the document, so let us see what is the simplest algorithm.

(Refer Slide Time: 11:17)



Unsupervised content selection; Luhn (1958)

Intuition

Choose sentences that have salient or informative words

Two approaches to define salient words

- *tf-idf:* weigh each word w_i in document j by tf-idf

$$weight(w_i) = tf_{ij} \times idf_i$$

- *Topic signatures:* choose a smaller set of salient words, specific to that domain

$$weight(w_i) = 1 \text{ if } w_i \text{ is a specific term (use mutual information)}$$

Weighing a sentence

$$weight(s) = \frac{1}{|S|} \sum_{w \in S} weight(w)$$

Pawan Goyal (IIT Kharagpur) Text Summarization - LexRank Week 11, Lecture 1 7 / 10

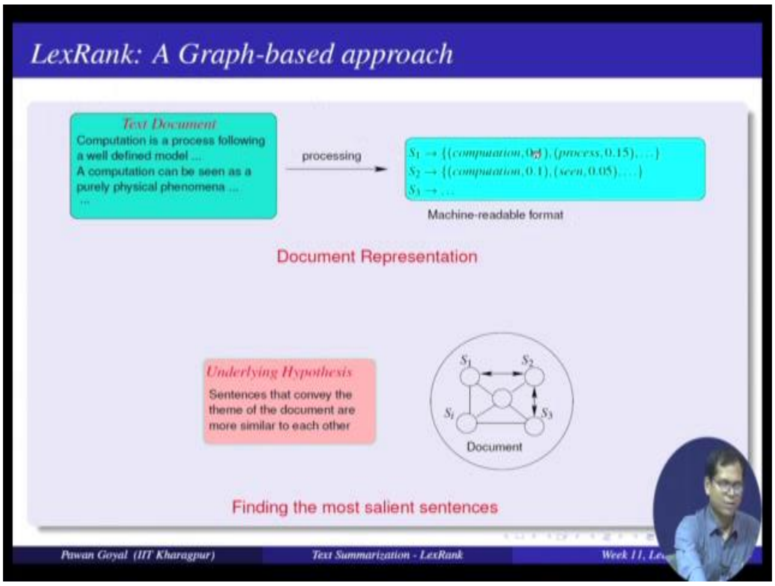
So, this was designed by Luhn in 1958; this is called unsupervised content selection, what is the idea few sentences that are very very informative and that how do you realize the sentences that are very informative. So, find the sentences that are having very salient or informative words; sentences that are having very informative words choose those. So, immediately the question is; for doing that I need to define what is the information contained for each of the word and given sentence I can add the information for each of the words.

So, how do I assign the information for each of the words, so what are the simple measures. So, one simple measure is you take a tf-idf for each word, so for each word in the document in the sentence, you find that what is the tf-idf of it. We already discussed how do we compute tf-idf of it and then take an average tf-idf for the whole sentence. Second approach would be, you define what are your topical words, what are your content this domain specific words and then take sentences that contain more of these topic bearing words. So, that is you define something like a topic signature and generally choose a smaller set of words that are specific to that domain and weight of word w i can be 1, if w i is a specific term in that domain and 0 otherwise. So, that is you are talking about politics, you know there are certain terms that are very important these terms into politics and they are called the topic signature for this domain and whichever word is in this topic signature is given a weight 1 otherwise 0.

So, now once you have defined a weight to each of your word using one of these methods, you find the weight of your sentence by taking adding the weight of each and every individual word and dividing by the length of the sentence; that is what is the average information of the words in the sentence, so you take the information for each of the word and take an average. This is a very very simple method, it does not see what the other sentences in the document and an anything else; it just assigns the information to each of the word independently and takes the important sentence by this method.

So, this is the simple method that you can use, but what is more much more principled method and much more common method is something called a LexRank method, this is coming from a this is like a graph based method for summarization.

(Refer Slide Time: 13:45)



LexRank: A Graph-based approach

Text Document
Computation is a process following a well defined model ...
A computation can be seen as a purely physical phenomena ...
...

processing

$S_1 \rightarrow \{(computation, 0.[illegible]), (process, 0.15), \ldots\}$
$S_2 \rightarrow \{(computation, 0.1), (seen, 0.05), \ldots\}$
$S_3 \rightarrow \ldots$

Machine-readable format

Document Representation

Underlying Hypothesis
Sentences that convey the theme of the document are more similar to each other

Finding the most salient sentences

Pawan Goyal (IIT Kharagpur)	Text Summarization - LexRank	Week 11, Le

So, what is that approach, so this flowchart kind of explains this whole approach; let us try to understand this. So, I have an input text document that contains a set of doc; set of sentences. So, like here you have a sentence like computation is a process following a well defined model and so on and second sentence a computation can be seen as a purely physical phenomena and so on so you have a set of sentences here.

So what is the first step, you find out what are the individual sentences here and provide their tf-idf weight. So, you say sentence S 1 contains the word computation with a tf-idf weight of 0.1, process with a tf-idf of 0.15 and so on; sentence S 2 contains computation with weight of 0.1 seen at the weight of 0.5 and so on and I will do it for all the sentences. This is like I am converting this document to a machine readable format, where I have different sentences; each sentence has some words with their different tf-idf weights.

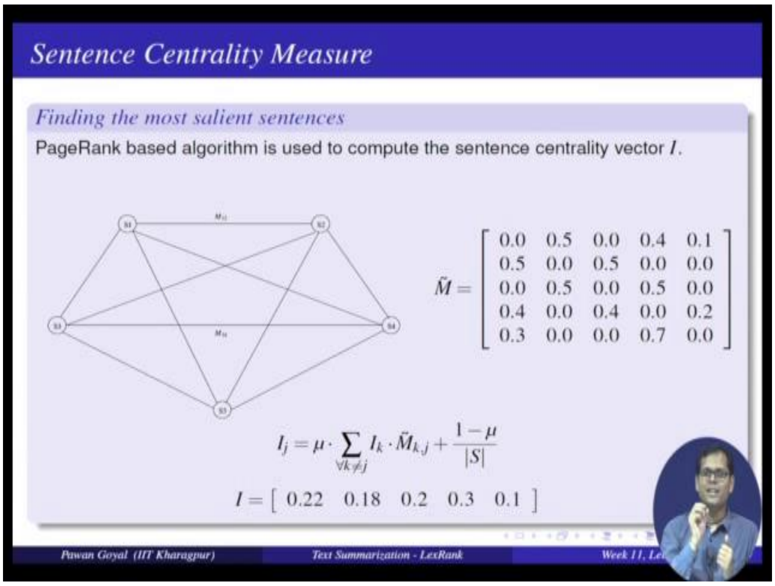
Now once I have done that, what is the next step? Now I treat it as a graph, where all these sentences are the nodes in the graph and edges depend on what is the cosine similarity of different sentences in this the pair of sentences. So, what I will do now; I will take all the sentences in my document and I will construct a graph where these are the nodes and the edges are what is similarity between every pair of sentence S 1 and S 2; how similar they are.

So, this graph is again very easily constructed because you already know what is the representation for each sentence; so you can come to the cosine similarity to get the similarity between two sentences. Now, once you have this graph then you use an hypothesis for getting the important sentences and what is the hypothesis that is the most important part here. So, the hypothesis says that when you are talking about the document, so it is a news document or research article; it is about the theme right; you have an underlying theme there.

Now sentence that are important to the document should be talking about that theme and because this theme is prevalent, this sentence should be similar to many other sentences in the document. So, that is a sentence that is similar to many other sentences should be called as the important or the relevant sentence and this is the main hypothesis that is used here and these other sentences should also be relevant. So, sentences that convey the theme of the document are more similar to each other.

So, now this can give us a nice intuition on how do we construct an algorithm; sentences that are talking about theme of a document are similar to each other. So, I am saying a document is important or a sentence is important; if many other important sentences are similar to it, it is similar to many other important sentences and that if you try to think of this hypothesis that is very close to what we do in the case of page rank algorithm. So, in page rank we say that a web page is important, if many important web pages link to that and here we are saying a sentence is important; if many important sentences are similar to that. That means, similarity we have to somewhat link it to incoming edges and so we directly convert that to some sort of a page rank algorithm, so what do we do there?

(Refer Slide Time: 17:19)



Sentence Centrality Measure

Finding the most salient sentences

PageRank based algorithm is used to compute the sentence centrality vector I.

$$\tilde{M} = \begin{bmatrix} 0.0 & 0.5 & 0.0 & 0.4 & 0.1 \\ 0.5 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.5 & 0.0 \\ 0.4 & 0.0 & 0.4 & 0.0 & 0.2 \\ 0.3 & 0.0 & 0.0 & 0.7 & 0.0 \end{bmatrix}$$

$$I_j = \mu \cdot \sum_{\forall k \neq j} I_k \cdot \tilde{M}_{k,j} + \frac{1-\mu}{|S|}$$

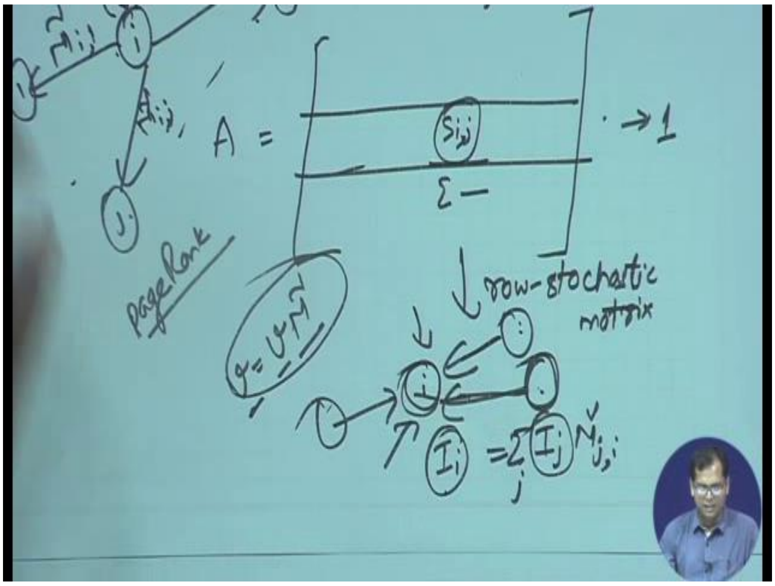
$$I = [\ 0.22 \quad 0.18 \quad 0.2 \quad 0.3 \quad 0.1\ ]$$

Pawan Goyal (IIT Kharagpur) Text Summarization - LexRank Week 11, Le

So, we construct a document graph with sentences as the vertices that is easy I know how many sentences are there that many vertices I construct. Then I compute the similarity between sentences and I construct some sort of a matrix. So, I compute similarity between S 1 and S 2, S 3 and S 4 and so on by cosine similarity and then I have a matrix format; of all the n cross, n pairs if there are n sentences in my document, I have a matrix of size n cross n that stores how similar each two sentences are.

Now, once I have constructed this matrix, I want to apply a page rank kind of algorithm to compute the informative score for each node. So, that is I want to compute the sentence centrality vector I and I has as many elements as the number of nodes. So, I will have a score for each of the node and the node that gets the highest informative score, has the highest page rank or is the one that is selected first. So, let me try to spend some time in discussing what is the rational for using this approach and how do you actually apply this algorithm.

(Refer Slide Time: 18:46)



So, what you are doing here, so you are constructing this matrix A; that is how similar two sentences i j. So, the element i j will tell me how similar two sentences i j are, so this is an initial matrix. So, now to be able to apply page rank algorithm; we need to convert that into a row stochastic matrix. So, convert that to a row stochastic matrix; what do I mean by row stochastic matrix; such that all the elements in an individual row add up to 1 and how do you do that? You will take all the elements sum them together; get a addition and divide it by the summation over all these elements and this will ensure that all the elements add up to 1, so summation of all these elements will add up to 1.

Now how does that help? So there we need to understand what is the intention for and this is called the M tilde here, this is my M tilde that is the row-stochastic matrix. Now what is the intention of adding this to 1, so adding to 1 gives me some sort of feel of probability right.

So, how we can see these integers probabilities, so now this M tilde i j will be the probability with which you are going from node i to node j. So, that is assume you are having a graph node i and there are all other nodes node 1, node 2, node j and node n up to node n. So, idea is that you are doing a random walk on this graph; at some point of time you are at node i; from node i you can go to all these nodes. So, there is a probability with which you can go to different nodes that is M tilde; i 1; M tilde i 2, M tilde i j; there can be some probability of staying here M tilde i i also; there is some probability with which you can go to different nodes and so you have distribution and you will sample some node and you will go to that and suppose you go to node j, then you start from there. So, this is a random walk that you do and when you do this random walk, it converges to something like v is equal to v M tilde, so v is the actual page rank vector that you are obtaining from there.

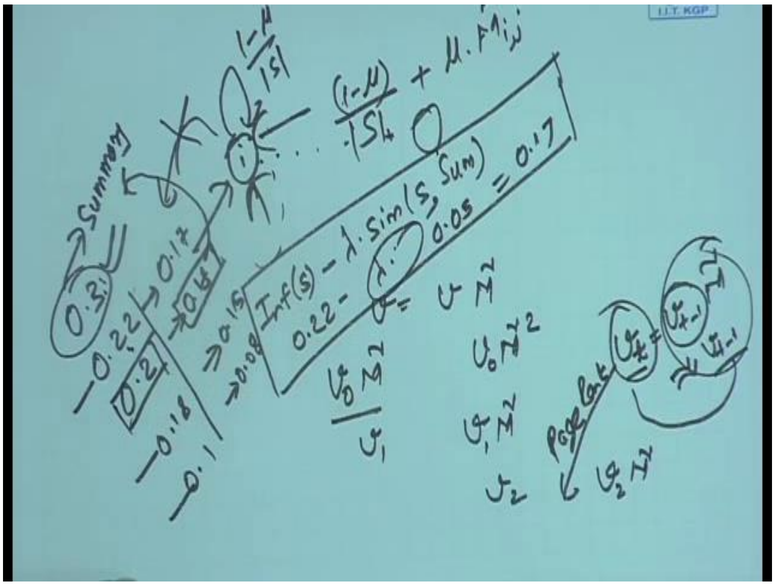
So, here in page rank what is important; a node, a webpage is important if many important web pages link to that. So, that is I have a node i; it is important if it is getting links from many other important pages. So, this is important, this is important, this is important and it is getting inlinks to many of these pages and how do they correlate. So, because you are getting inlinks from many important pages what does that mean?

So, the translation between this random walk and this page rank value is the high page rank value indicates that, if you do a random walk for sufficiently long number of time; you will stay on that node for a larger fraction of time. So, suppose this node has a higher page rank than this; that means, if you do a random walk for large number of time; this node, you will stay at this node for longer time than at this node. So, now if this node is linking to this node; lot of lot many times what does that mean because you will be staying at this node for longer times and from here you have a good probability of reaching to this node. So, accordingly you will also stay at this node for a longer time and if you have it from many important nodes; that means, there is a high probability that you will be staying at this node for a longer number of times.

So, that is if a node gets incoming edges from many important nodes then it has a high page rank and that is formulated by this simple equation. So, that is if I want to write it as a equation; I will say importance of this node i would be importance of node i will be if I write crudely I will say will be proportional to importance of all the other nodes j, that are linking to it and the way the probability with which they are linking to it times M tilde j i. So, i j times M tilde j i and I will sum over all j and you will see that this will actually take it to this formula v is equal to v M tilde and that is my page rank algorithm; importance of a node is; importance of all other nodes that are giving incoming link times the probability of going from that node to this node; how high this incoming linkage.

So, and then you iteratively compute these i; all these i values and there is a very simple formula for computing these iteratively and what is that.

(Refer Slide Time: 24:26)



So, we know the equation is v is equal to v M tilde, so how do you compute it; you start with some initial v 0 and keep on multiplying with m tilde. So, v 0 M tilde is your v 1 then you got v 1 m tilde, so this will be like v naught, M tilde square then you got this is your v 2, then you go v 2; M tilde, keep on doing that. Now what will happen because this is my equation, it will converge at some time where v t that is v t minus 1 M tilde is very similar to v t minus 1 so; that means, now v t and v t minus 1 are roughly same and this is your actual equation.

So, it will converge at some time, so you keep on multiplying by M tilde at some point of time and at every time you keep on finding the difference; v t minus 1 minus v t find out the difference; at some point of time this will be below threshold at that time you will stop and whatever v t minus 1 or v t you get at this point; this becomes your page rank and that is how you compute your page rank.

So, now we saw what is the rationale of using page rank and how do we compute page rank, but now if you go to the actual formula; this looks slightly more complicated than what we were saying. Now this is similar right i j is equal to summation over i k times M tilde k j importance of k and the probability of going from k to j summation over i k, but you see something like a mu and 1 minus mu divided by S. So, let me see what is this mu and S; S is the number of sentences here and mu is something like a small teleport probability what is the idea?

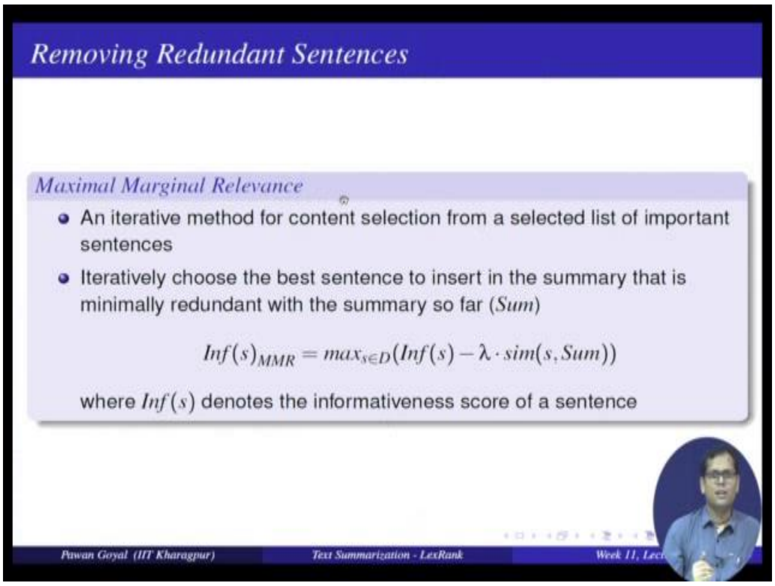
Suppose in your whole graph; there is a node i that has some incoming edges, but there is no outgoing edge from here, so there is no outgoing edge. So, when you are doing a random walk; once you come to node i, you cannot go back; you cannot go to any other node in the graph and your random walk is stuck and this creates a problem in defining the statistic probabilities and all that. So, that is why you say from each node I will assign a small probability to all the nodes.

So, this probability is nothing, but 1 minus mu, we equally distribute among all the S nodes divided by S. So, there is a probability 1 minus mu divided by S going to all the nodes, they will always have self probability 1 minus mu divided by S and to balance the probabilities; you multiply plus mu and the actual probability M tilde i j. Suppose this was M tilde i j initially, now the probability becomes 1 minus mu divided by S plus mu M tilde i j and you will see that now the probability will still add up to 1 and that is the only thing that you are seeing here, so i j is mu times this plus 1 minus mu divide by length of the (Refer Time: 27:55).

So, now if you take in general the value of mu that you will take here is roughly close to 0.85 that is a common general value. So, if we take mu is equal to 0.85 and try to run this algorithm, so I will encourage that you try to run this algorithm, you will find that these are the page rank values that you will get for these five sentences. Now how would you pick a sentence from here, you will take the one that is having the highest score. So, we will take the sentence S 4 first, then if you want a longer summary then you will take S 1 further longer then S 3; more than that will be it will not be a summary. So, we will take S 4 then S 1 and S 3 if required, so this is how you will construct a summary.

That is the first part of content selection, how do we select the important sentence from the document. Now there will be one problem here that is suppose I select the S 4 sentence, but S 1 is quite similar to this, but this is coming how to be irrelevant as per this approach. So, I want to give some less weight to a sentence if it is similar to some already existing sentence in the summary and that is why I can do some sort of diversity and I will quickly tell what will be the approach we will use for diversity and this is called the maximum marginal relevance.

(Refer Slide Time: 29:19)



Removing Redundant Sentences

Maximal Marginal Relevance

- An iterative method for content selection from a selected list of important sentences
- Iteratively choose the best sentence to insert in the summary that is minimally redundant with the summary so far (Sum)

$$Inf(s)_{MMR} = max_{s \in D}(Inf(s) - \lambda \cdot sim(s, Sum))$$

where $Inf(s)$ denotes the informativeness score of a sentence

Pawan Goyal (IIT Kharagpur) Text Summarization - LexRank Week 11, Lec

So, given something is already in the summary; how much more relevance you are bringing in by the new sentence. So, this is like an iterative method for content selection from a selected list of important sentences. So, what is the idea; you iteratively choose the best sentence to insert in the summary that is minimally redundant with the summary so far. So, informativeness of the sentence will be among all the sentences in the document; informativeness of the sentence minus lambda times similarity of sentence with the already existing summary and whichever gives the highest informativeness is included in the summary, so what is the idea?

So, suppose you have the sentences with this informativeness 0.3, 0.22, 0.2, 0.18 and say 0.1 informativeness of five different sentences. So, you took this sentence for your summary; summary contains this sentence one. Now as per the relevance, you will take sentence S 2, but what this approach says; now you again re rank them, what is the criteria; this would be informative score of a sentence S minus lambda times similarity of S with the summary. So, that will be 0.22 minus lambda times similarity of this sentence with this sentence; suppose it comes out to be lambda times similarity this whole thing comes out to be 0.05, so this becomes 0.17. You will do that for all the sentences, suppose it comes out to be 0.17; this comes out to be 0.18, this comes out to be 0.15, this is 0.08. So, then you will again sort them and take the one with the highest value as per MMR and this goes to your summary.

Now your summary has two sentences; again you will compute this informativeness minus lambda times similarity of this sentence with all the sentences in your summary. You will again rank this, this and this and you will try to take the one that is having the highest score as per MMR and you will do that iteratively until you have got the desired length of your summary and that is the idea of taking redundancy.

So, you can see that how we are helping, so initially I would have taken this sentence followed by this sentence, but this sentence is very similar to this sentence. So, this approach helped me to choose another sentence that is not so close to the original sentence, so you are also having the diversity, so this was one approach for doing summarization. So, in the next lecture we will talk about another approach for doing summarization that is based on some sort of optimization. So, see you then.