FOUNDATION OF DIGITAL BUSINESS

Surojit Mookherjee

Vinod Gupta School of Management

Indian Institute of Technology Kharagpur

Week 07

Lecture 32

Lecture 32 : NLP and Computer Vision Applications

Good morning. Continuing with my module number 9, which is about tools and techniques for AI. Today, in this session, I will be touching upon natural language processing and computer vision applications for business problems. So, for NLP applications, it has now become very popular and is finding applications in a host of areas. And some of these are listed here for your reference text classification, conversational agents, translation, sentiment analysis, text summarization, information retrieval, and information extraction.

Apart from these, there could be some more. So, this is not an exhaustive list, just some of the few important areas where it is now being used extensively. So, if you want to define what NLP is, it is any method that processes, analyzes, and retrieves text data. So, anything to do with text. So, text classification—what is it doing here? Document categorization is the automatic labeling of documents and text units into known categories.

For example, a use case for manual classification can be very time-consuming, error-prone, and cost-prohibitive, especially for those working in service management areas. We are doing customer service, and there are various sorts of customer issues that come in. a software system, you have a tool and lot of complaints or whatever their issues they keep coming and you as a service agent you have to team or call centers, you have to reassign those tickets to the relevant person in your business. So, it could be a refund related issue, it could be a complaint, it could be a login type of technical issue that customer is not able to login

say, for net banking or any such operations. So, you have to go on assigning. So, that is a very manual job, which is very common in the service industry. Now that is being replaced gradually by NLP tools for text classification. So, the tool has been trained to identify classify these tickets into.

certain categories. These categories are preset predefined and you are training the tool when you are training you are using training data of course, the old previous all the problems which are there in your database or repository. So, you are using those data sets to train the model and teaching them which problem is in which category. So, something should go to a refund category, certain problems should go to the complaint category, certain category should go to the login category.

It is a very simple use case. So, they are turning to machine learning and natural language processing to automatically organizing the text into one or several of such predefined categories. LinkedIn for example, I am sure many of you have heard about using LinkedIn, they use text classification techniques to flag profiles that contain say inappropriate content. which can range from profanity to advertisements for illegal services. So, you can post anything in LinkedIn like the tool does not stop you from posting anything, but how do that LinkedIn tool ensure that whatever is

being posted is proper legal does not hurt anybody else sentiment etcetera. So, only option is to go for such. NLP tools because otherwise doing it manually means you have to employ a huge number of people who will be just sitting and going through various LinkedIn posts that we an impossible task given the size of their scale and operations. Practical text classification examples.

Some examples is Facebook is famous for whatever it is and there are lot of hate speeches go lot of feedbacks are given, you post something even in YouTube or many such other social media many people give very negative feedbacks, not from negative from the context perspective, but just general hate or something to do with religion, something to do with race, country whatever. So, some of the examples here.

here is mentioned here I will not probably read out you can read it etcetera you can see religious hate, then there is some generalized hate, why do some fill the wet bag very improper language something which we refer to commonly as unparliamentary language which are being used and these need to be weeded out, these need to be removed. than violent or explicit content. So, sometimes this hate term what is actually hate spreading hate that becomes

slightly more difficult compared to very explicit content. So, you can openly use very bad words or call out a religion or race by name. So, those could be identified easily, but it could be a descriptive head might be difficult to classify or identify, but then that is where you use AI tools if you properly train them the NLP tools can do this job quite well. Your Gmail you have a spam folder where it detects any mail which it thinks that it should not come to your inbox, it automatically transfers it to the spam folder.

And you can of course, go and check that mail and find out whether it was a genuine spam or not a spam and you can then recover the mail. So, again there they are using an NLP tool for that purpose. And something like a net promoter score, so if you get review feedbacks. So, you have some negative feedbacks, you have some neutral feedbacks and then you have some positive feedbacks which are promoting your whatever product or service. So, the tool can calculate something like a net promoter score.

So, you can classify them in this category detractors, passives and promoters and then you say what is my net promoter score is the promoter minus the detractors. So, you can get you do not have to do that manually all sorts every review comment will be assessed by the NLP tool classified in these three categories and then do the simple formula to tell you what is your score or what is your CSAT index for example, customer satisfaction index. So, all of these cases we can use where we have a feedback system or a database which is accessible by say thousands of customers they can give the feedback and then I use this NLP tool to classify them in such categories

which I want I define the categories and then get by whatever I want to get out of that. Conversational agents that is what we have most of us are I am sure familiar with. They communicate with users in natural language with text, speech or both. They can come in two categories one is the virtual assistants and other is the chat bot. Virtual assistants are also referred to sometimes digital assistants or AI assistants and they are designed to complete specific tasks

and are said to have reasonably short conversation with users like Alexa for example. Siri or Google assistant that examples of this digital or AI assistants or virtual assistants whatever you want to call them. They understand human commands in our natural language and can complete tasks like setting an appointment in your calendar, calling a friend, finding restaurants, giving driving directions and switching on your TV. People are using these agents on their websites to answer customer questions.

and resolve simple customer issues. So, they can be a logical virtual extension you can think of that as a chatbot as well. But here normally we are used to these agents as physical devices hardware stuff which we keep in our home or office or wherever. So, Alexa device or a CD or a Google assistant and then it does whatever we ask it to do. It could play a song which you like.

give you weather information, give you a news, dial a number and things like that. But you have to be careful about this AI tools in the sense it has been reported that they are also listening to you. So, it is a two way communication. So, not only they are doing what you are asking them, but possibly they are also listening to you. But if you are having a discussion meeting or even your house your domestic discussion or fight argument whatever.

So, it is probably listening and why it is meant to do that? One of the purpose is that all of these everybody wants to know what you want. like to do or what you like to buy or what you like to eat. So, what are your likes the whole world the whole marketing world wants to know what all of us want like. So, that this information then can be passed on or sold it can be monetized it can be sold to people who are interested like advertisers would like to know what

restaurants food I like, what restaurants I like to visit or what clothes I like to wear or what color I like etcetera. So, in other words the flip side the negative side of this is that there is no privacy in the real sense of the word. So, if a device which is very useful to me for various purposes can also pass information it is listening eavesdropping kind of thing acting like a spy also you can say something like that. So, all of these things which are happening and the technology is being used to find out my whatever likes dislikes wishes then

which information in future in later terms can be used by an ad agency to exploit my likes and dislikes. Chatbots on the other hand are designed to have extended conversation with people, it mimics chats in human to human conversations rather than focusing on a particular topic. So, chatbots are that way one way. So, whatever you type it will respond to that it does not normally it does not have audio response. So, if it is a voice responding system then probably it also listens to your voice.

So, it takes a voice command then it can of course, listen eavesdrop like if two people are chatting in front of the chat box is open probably theoretically yes it can record whatever is being discussed and then that could be analyzed by an AI tool at the back end by

whosoever is interested do so. So, when it comes to conversational agents what is more relevant to business is that AI assistance that can work on tasks without needing human intervention. So, this is what the future is or is already there, but people are putting it more and more into use and we are calling it agentification that not only it will tell me something it will also do certain actions follow up actions it will complete the task.

So, I give an instruction buy a plane ticket, find out the best flight available cheapest flight or whatever etcetera from here to here and buy the ticket. So, it not only finds out which is the cheapest flight available, but also books your ticket according to whatever preferences you have given time zone time etcetera you want a morning flight evening flight on a particular day etcetera. Machine translation. There is an example which is an example of a incorrect translation.

So, there is a French idiom to English translation using Google translate. The input is an idiom French which means he is getting on my nerves, but if you see that actual translation it has done he runs to me on the beam. Now this is not uncommon. But the good news is that machine translation is handy for simple translation tasks in business application and the quality has definitely much improved over the time because AI is now being used extensively

that is LNP is being used extensively and it is much more developed model and it can now operate in multiple languages. So, from English you can probably translate to any of the Indian languages. So, it can render web content in different language depending on the visitors language settings. So, this can become very useful proof to very useful in say government setups in countries like India where you are using multiple languages. So, in any state they can use English input and get it translated into whatever local language they want to.

So, this is giving tremendous strength to government departments who are into public service and government most of the departments are into public service. So, the language is no longer a barrier literally. Even Google translate if you are visiting a country like say Japan and you can just speak or you can take a picture of whatever is written in Japanese and it will translate you to English for example. So, things are becoming much more convenient for say international or global travelers.

We are visiting multiple countries like Europe language is a problem everybody in all countries most countries they do not speak English, Japan, China etcetera all of these countries English is a problem need such tools. Translating customer support requests

that are in a different language from the supported native language etc. And then standardized datasets that are in different language before they are used for any downstream analysis. So, there are various use cases for translation application of NLP.

Sentiment analysis is being used in a very big way in marketing domain, because people want to know from the feedback what the customer is actually feeling. So, they may not express it very clearly. So, you have to make out, read out, interpret whatever they written the text and do a sentiment analysis using NLP tools to find out what actually they are meaning. Then automatic interpretation and summarization of emotions within the text data. So, I want to extract that emotion.

For example, in predicting emotions in tweets, emotions can be positive, negative or neutral. It can also be more granular when you detect elements such as anger, joy, sadness and disgust. So, this sentiment analysis enables business to analyze the customer sentiment towards your brand towards your product and the services that you are delivering etcetera using online conversations or direct feedback. So, all of us most of us we give feedbacks.

So, when we travel and they want also request for feedback we give feedbacks any service may be hospital service, restaurant service or a hotel service or you are travelling somewhere the tourist service. So, all of these services they request for feedback. And we do give feedback, text feedback and then these tools are being used by the marketers to analyze your the emotion behind your feedback. It analyzes the business to analyze your enables business to analyze your customer sentiment towards all these products, bands, services etcetera.

And with these companies can better understand the company's likes and dislikes and then accordingly change. and improve and find out ways of innovating things so that your negative whatever sentiment you expressed next time or other customers they will have much more better feedback. Summarizing text, this is again a very powerful application or use case for NLP. It was automatically reading some text content and generating a summary. So, you give a 500 page document or a 200 page document and ask for it.

two page summary or a three page summary. So, it can be so productive to people who are working in the say the legal services we have to go through read through hundreds and pages of legal references case history etcetera, but then your boss your senior wants a one or two page summary gist of the whole thing. So, doing it manually you can understand it is very time taking time consuming, but if you can use such tools it

becomes very fast. So, it is concise and readable, it picks up on aspects naturally, contains important details and displays how many people said it, how many times it was mentioned the count of it.

So, in the bracket you can see there is giving the count of people who has mentioned that any of these aspects say the interior is good. or the build quality is good etcetera, the DVD surround sound is amazing etcetera. So, not only get the important essence of the feedback, but also the number of people who have made such comments. So, summarizing free text survey comments example Amazon's read reviews. that mentioned etcetera.

So, in Amazon there you get review after review after review. So, you can it can summarize all those reviews to give you the gist. So, these are being used in research extensively used in research in management research for doing sentiment analysis from published review comments in various e-commerce and any other online sites. Summarization can be of two types. extractive and abstractive.

Extractive where it actually just picks up portions of that document by selecting combining the key passage and they combine those selected passages and come out with the summary statement. So, they will focus on extracting the most important sentences without creating a new content. So, just picking up the main key points, key sentences, actual sentences they will not change it or add any new content. The goal is to preserve the meaning of the original text while condensing it, you are basically condensing it.

But an abstractive summarization will generate entirely new sentence to convey the key ideas. So, it will try to understand the context and then write new sentences, so that that context is properly expressed. So, unlike the extractive one, it selects and rearranging sentences just mechanically. Abstractive will rephrase the information in a more concise and coherent manner, often using new vocabulary that was not present in original text. So, it will write new words or sentences.

So, it has gained prominence with the advent of transformer models which have revolutionized NLP tasks. So, you heard about the transformer model, there it does what it does it is predicts the new word. So, it goes on adding new word after word after word to form new sentences, new paragraph, new totally new text. It is a unique architecture that will significantly improve the performance. So, what we want really is given a text or a volume of data textual data.

I want to extract or summarize the essential meaning or convey the essential context which it is trying to say. If I read a 500 page document, then I need to know what is there the most important points covered in that document. So, that is where abstractive summarization comes into use, and extractive summarization will not be of much help. So, this is obviously much more powerful. Information retrieval is about finding documents that satisfy a user's need from a large pool of documents.

Google search is a classic example of an IR system; it gets you the information that you need from the entire web. So, it started maybe in the mid-90s—at that time, there was no AI, so to say, but now they have introduced an AI engine. So, the Google search has become much more efficient. Information extraction is the process of pulling out specific content from the text. So, it is a powerful tool when you want to precisely extract the content buried within large blocks of text and images.

It is a kind of extension of summarization, but here you specifically say, 'Take out this particular information,' whereas in summarization, you abstractly summarize. It brings out whatever it understands—the tool which identifies what is important after going through the entire text. But here, you specifically say, 'Within this text, wherever it is describing or addressing something, bring that out.' For example, extracting appointment information from your emails automates the process of adding appointments to your calendar. So, let's say you are a practicing doctor, and people send you appointment requests through emails.

So, you can use such tools to extract from the emails those referring to an appointment— seeking a date and time—and accordingly, you put them in your calendar. So, the tool does it; it puts it in your calendar. You can use Gmail for that purpose. And Google does this by reading the emails say about a flight confirmation or a concert and offers to add those events to your calendar. So, any flights you have booked and then your confirmation comes to your mail your PNR, date of flight or if there is any rescheduling of the flight etcetera information they normally the airlines they send by mail.

So, you can use this tool to extract that particular information and put it in your calendar the flight is on such and such date or such and such time etcetera. The various models of test generation because that is now one of the most common use case for NLP. So, tools like GPT-4, ChatGPT, etcetera is very familiar with the generate text by predicting one word at a time based on the sequence of words that came before it. This is the transformer model which I was talking about sometime back. It ensures that the generated text

follows a logical and coherent flow much like how humans write by thinking about the next word based on the context, the previous context.

So, the same word can be used to have applications in multiple contexts. So, this is what the whole transformer model is there is a classic Google paper of dated 2017 those who are interested with technical obviously, it is quite technical you can read that paper. I mentioned that in the reference. There is set to set models these are commonly used in tasks like machine translation where an input sentence as a whole in a language is transformed into an output sequence a translated sentence in another language.

So, this is one reason it is not going word by word by word or going by the context of it. Hence, sometimes those like the example I showed you earlier some of the translations do not make much sense. fine tuned models, pretend AI models that can be further customized. We have been talking about this formulation models and then fine tuning them for some specific purpose. So, here like you can using fine tune them for using specific datasets to specialize in domains such as generating medical reports, generating legal documents or financial summaries.

So, you can use that same text generation tool by fine tuning your formulation model. Retrieval augmented generation RAG, this is again you must be hearing it, this is now becoming very popular what is RAG? It consists of three parts retrieve to find out the information, augment, add it to AI's own knowledge and generate a better create a better response. So, it is a very interesting tool, it will first search external sources for relevant information based on the users query of relying only on existing It just does not rely only on the data which it is trained on if you give a query it will first go out to the net and find out latest of that in that subject then

combine that with whatever it has been trained on and with this combination it comes up gives you the output the result generates the output which is of course, an updated version of whatever it could do had it not checked the current information available in the net. So, external data is referring to the new information beyond the LLMs original training data set. So, we are using that foundation model.

So, we have been talking about such the foundation models like when chat GPT was developed it was trained till data available as of November 2021 they announced and then they release the products later much later sometimes later. So, whatever it was giving an output was based on whatever training data has been provided to the tool as of

November 2021. So, any data for December 2021 or beyond was not used for the chaining.

So, the model does not know about the data, but here that is the solution. So, it can search various other datasets and various other sources such as API databases or document repositories and may exist in different formats like of course, multimodal text files or structure records. And then to make this data understandable to AI it is divided into chunks in cases of massive datasets and converted into numerical representations embeddings. So, this is how the model works because everything has to be numerical that is something like

And then you convert that into call that as embeddings using specialized models and then stored in a vector database. So, vectorization has to happen because anything because after end of the day a computer has to work on numbers and not text. So, any text is reads needs to be converted to a number. So, here in this particular process technique it is called as embedding and then it creates a knowledge library that the AI system can reference during retrieval. So, from the external source you collect all your text informations, vectorize them of course, and you are calling them as embeddings and then you basically creating a virtual knowledge library.

So, when you submit a query what happens this retrieving of the relevant information, the system will convert into a vector representation and matches it against the stored vectors in that database. So, this enables precise retrieval of the most relevant information. So, one is whatever you have trained on and then you now create an additional vector database which you are calling is a library. So, your query you will first see that and then see whatever it is available within trained itself and then combine the two and give you the final output. Now what problems does it solve?

Why it is becoming so popular? Factual inaccuracies and hallucination. So, the error will gets minimized or reduced from as compared to traditional generative models, because it is retrieving verified external data. So, your response will become much more refined. Outdated information obviously, so you are not relying on a static model, but you are going dynamically relieving up to date information.

ensuring that relevance is accurate in real time. Contextual relevance, so often they struggle with maintaining the context in complex or multi turn conversations, but since it is relieving relevant documents from that again to enrich the context hence improving the coherence and relevance. Domain specific knowledge again the generic model the

foundation model may lack expertise in specialized fields, but the RAG will integrate domain specific external knowledge for tailored and precise responses.

Cost and efficiency fine tuning large models for specific task is expensive. So, RAG eliminates the need for deep training by dynamically relieving relevant data. So, you do not have to train that training is very expensive. So, I take a foundation model and then I use this RAG model because that once I given query it will immediately go out and search for relevant information, collect it, summarize it, link it up with whatever it has been trained upon and then combine and giving me the output.

So, I do not need to go on every time training at the foundation model into domain specific things. Scalability across domains, it is adaptable to diverse industries and health from healthcare to finance without extensive retraining and it is highly scalable. Some business applications of computer vision AI tools. Now moving to this computer vision AI tools, it is commonly used in various applications. I have listed down these applications retail and e-commerce, inventory management, visual search, virtual try ons for your dresses, your fashion industry, etc.

online purchase. In manufacturing quality control, predictive maintenance, robotics and automation, healthcare medical imaging, we have been discussing this earlier, imaging trying to predict the model whether you have a particular disease or not, surgical assistance for robotic surgeries, and patient monitoring. In automotive and transportation, autonomous vehicles, of course, because they use all those camera-based sensors to detect whatever is on the road. Driver monitoring: you can monitor the driver, even the manual driver, then it can use computer vision systems to test the alertness of the driver, whether the driver is falling asleep.

And traffic management all of us know because in most of the big cities in the traffic cameras to find out whether you have crossed the line, broken the signal, or are speeding, or things like that. Security and surveillance: facial recognition, anomaly detection, crowd monitoring. Facial recognition—all of us are using it heavily in airports, in DigiYatra. Agriculture crop monitoring, livestock management, precision farming everywhere you need something to see visual thing you are replacing human beings with cameras for remote access, and you are getting much more volumes of data because it is 24/7.

Finance and insurance: you can also do document verification, damage assessment, and fraud detection. In logistics and supply chain: package inspection, warehouse automation,

and route optimization. And in construction and real estate: site monitoring properties. You can see in all these various multiple domains, you have numerous applications using computer vision tools. Use cases for some open-source computer vision tools. Like these are the open-source tools: OpenCV, TensorFlow, PyTorch, etc.

So, you can use them for various image and video processing, object detection, etc. So, I have just mentioned some common use cases for your reference. So, with that, I will end today's session on NLP and computer vision. Thank you very much.