ALE | Where Everything Connects

Artificial intelligence applied to enterprise communications

Alcatel·Lucent Enterprise

Table of contents

Abstract

Being exposed to all the information flows around artificial intelligence (AI) and machine learning (ML) initiatives, the leaders responsible for establishing and evolving enterprise communications practices often wonder what AI and ML are and how these novel technologies can be applied in their area of responsibility. This white paper explores the concepts of artificial intelligence and machine learning, discusses the different types of information generated by enterprise communications solutions and the potential business impact from applying AI and ML techniques to this information.

What is artificial intelligence? What is machine learning?

Invented in the middle of the 20th century, the term "artificial intelligence" means a technology that is capable of performing tasks that are characteristic of human intelligence. Such tasks may include problem solving, learning new things, recognizing objects and sounds, understanding human language and so on. This definition of AI is rather general and while interesting in philosophical discussions, not very useful for engineering. In a practical sense, AI can be considered to be many un-related approaches to problems where each approach is an extremely good application when applied to a specific task, for example, recognizing given objects in photos.

The term "machine learning" was coined not long after artificial intelligence and it means the ability to learn without being explicitly programmed. As you can see, machine learning is a way to provide the machine with AI capabilities. Theoretically, AI can be constructed without ML, but this would require building programs based on complex rules and decision-trees. Instead of hard coded software routines with specific instructions to accomplish a specific task, machine learning is a way to train an algorithm so that it can learn information, necessary for accomplishing this task. Training means providing the algorithm with a lot of learning data which allows the algorithm to deduce necessary logic. The ability to learn is usually defined as a capability to improve the performance of task execution based on the quantity and the quality of the data, used for the training. Machine learning is often based on computational statistics and mathematical optimization techniques.

In general, there are three types of machine learning:

Supervised learning

This type of learning uses correct answers as the primary means for training. It is applied to classification problems when the program needs to correctly identify some objects as belonging to a given class.

With this technique, a human feeds the program with input data, such as pictures with different objects, and providing correct answers, such as revealing which pictures contain cats. Supervised learning algorithms try to understand dependencies between what they see in the pictures (fur, round eyes, etc.), and the label "cat". After the learning phase is finished, the algorithms are exploited and used to identify cats on presented unlabeled pictures. The functioning of the algorithms is based on the preliminary knowledge that a picture must contain fur, round eyes, and more to be classified as a picture with a cat.

Unsupervised learning

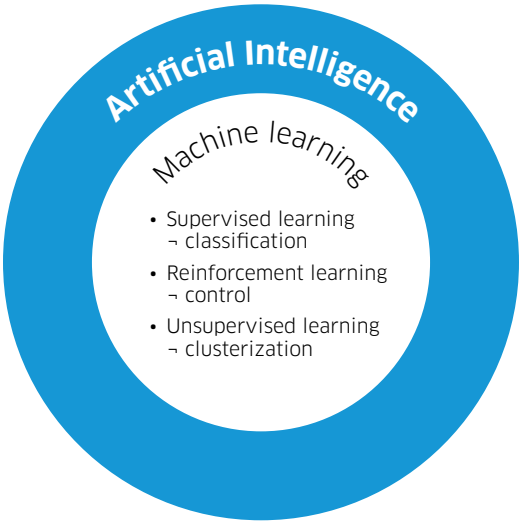
In this type of learning there is no teacher. The program tries to deduce information from the samples presented. Unsupervised learning is used for the clustering problem, when input data is attributed to a group, formed automatically by a program and not proposed by a human. These algorithms are based on found patterns, discovered automatically by their specificity compared to all other data.

Given a set of images, some of which depict cats, unsupervised learning will group the images with the cats in one group, based on common characteristics of the images, while not knowing that these are cats being depicted.

Reinforcement learning

This type of learning is applied to the tasks where artificial intelligence must make decisions affecting the environment in which it functions. Examples of such systems are self-driving cars and chess playing computers. The nature of reinforcement learning is that it uses observations gathered from the interaction with the environment to take actions that would maximize the reward or minimize the risk. This process is repeated iteratively, resulting in better results as time passes. In the human world, reinforcement learning is known as the trial and error approach.

This graphic summarizes these approaches.



All three types of machine learning are under active development by researchers and the industry, and applied in different areas of enterprise processes.

Machine learning in enterprise communications

Let's consider how artificial intelligence and machine learning can be used in the enterprise communications domain. Since we are interested mostly in automatic logic to minimize manual efforts, we will only address machine learning as the most prominent part of artificial intelligence.

When discussing ML in enterprise communications context we can distinguish two levels of its use: System and application.

System level means using ML with system generated information from different types of system logs. System level doesn't add any new functionality that is visible to an end user, but it improves the quality of the audio/video streams and overall system sustainability.

Application level implies the end-user features that can be developed or enhanced by application of ML techniques. These techniques are applied to user generated information, which represents the content of the communications: chat messages, conference transcripts, and different forms of context.

In this paper we are investigating the following types of information:

- Audio and video streams
- Application and network logs
- Conversation content
- General company context

Audio and video streams

Audio and video streams can be processed by ML to provide both system and application level benefits.

System level

The audio and video content is transferred on the network in a form generated by the codecs. The codecs are complicated creatures as their function is regulated by many different parameters. These parameters include bit rate, packetization size, buffers, and timeouts. Deciding which ones to use is not a trivial task as it depends on user preferences and network conditions. Both user preferences and network conditions can change dynamically in run-time. Provided with the historical information on how the given codec parameters affected user quality under the given network conditions, ML can deduce which parameters would be the most beneficial under current network conditions.

Codec parameter values can be suggested by ML not only for an individual user, but for a group of users as well. If several people are located behind a shared network link, ML analyzes the streams together and applies appropriate settings to each stream in the set. Whether it's better to send every participant's video stream with mediocre quality or only the currently active speaker with better quality, ML can make a choice based on previous feedback of the participants.

Going further, by applying ML we understand which parts of the frame are more important and encode that particular part with higher quality than other parts. For example, in a video conference the most important part is the talking head that is usually in the center of the frame. This part is dynamic compared to the static parts representing a conference room. This fact is used by ML to understand which part of the frame should be encoded with better quality.

Another interesting application of ML is in regeneration of video frame information to minimize network bandwidth requirements. The stream is sent with the lowest quality level, and the client restores it as a high-quality picture by applying ML. In this case, ML is pretrained on a set of different pictures and then info from this training is used to increase the quality of video stream frames.

Audio streams can be analyzed for voice quality augmentation by noise suppression. Instead of using a well-known approach based on the signal/noise ratio, ML understands the content of the speech on a semantic level and uses this information to separate the noise from the content.

Application level

Application level features, provided by ML, are defined by what you can hear and what you can see from the stream.

First, the audio content is transcribed into the text, which is called ASR (Automatic Speech Recognition). This gives a lot of useful functionality which is described in the next chapter.

One special case of ASR is when you are talking to the system itself, not to people. This functionality is called voice assistant. The meaning of the phrases that you pronounce is caught by the system and used as direct instructions.

Next the video content of the stream is analyzed. Here ML helps identify the participants of a meeting if image recognition techniques are combined with a database of employee headshots. Also, image recognition permits marking the words of each speaker with the correct name in the meeting transcript.

Next it attempts to understand the nonverbal communication between participants. If a participant looks at another one and proposes a given action point, and the second person nods, then ML is able to recognize this and assign the second person to this action point.

The next set of features is auto-framing, which mimics the work of a human operator with panning, zooming and tilting of the camera. Thanks to ML applied to the video stream, these actions can be automatically executed to improve the overall user experience. ML helps as well to select the most appropriate camera, if there are several, and de-skew an image, for example, of an angled whiteboard.

The TTS (Text-To-Speech) transformation supported by ML is useful for when it's desirable to have the result vocalized by a given person's voice, but the person is unavailable to record necessary samples. The available samples can be processed by ML to extract unique information inherent to the voice, then this information is used to vocalize any messages.

Interesting functionality can be achieved by extracting useful signals from the surrounding "noise." For audio, it means that the system separates the voice of a speaker from other sounds, like loud typing, heavy breathing, dogs barking. Elimination of such background sounds can significantly improve the quality of the end-user's experience during a conference. In this case, ML can use a semantic model of the speech to better understand what was said, and not be based only on traditional signal/noise ratio.

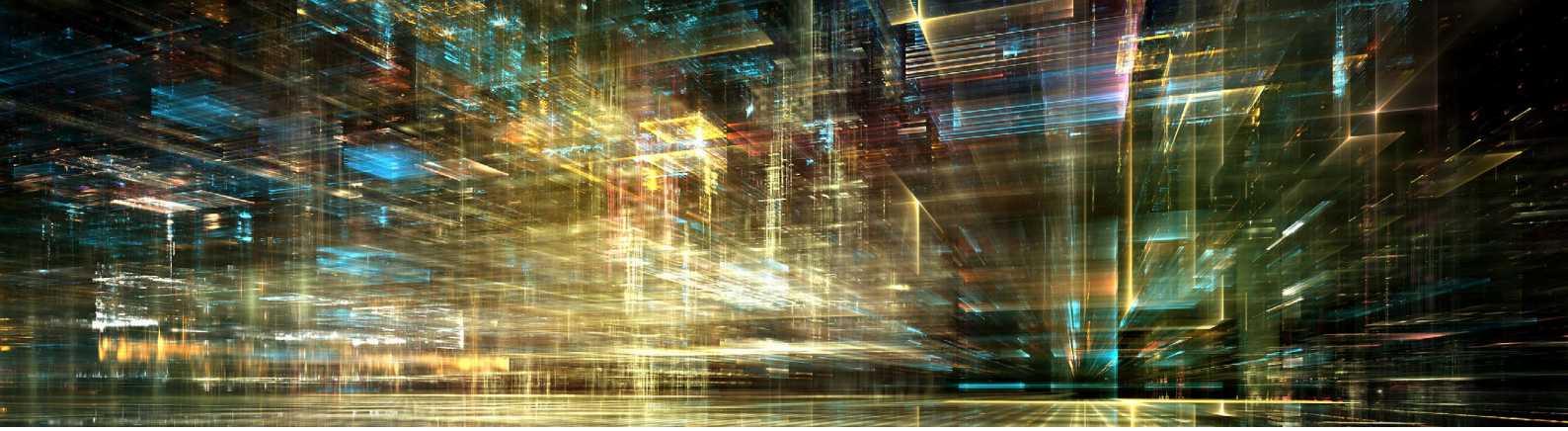
When applied to video, extracting a speaker's image is possible to completely replace the background. It is also possible to blur the background to ensure privacy. For example, not showing sensitive information that may be written on a whiteboard behind the speaker.

Application and network logs

System level

Being applied at system level, ML can be used as an essential tool against anomalies and external attacks. It continuously watches over system activities that are recorded in the application and network logs, detects anomalies and attacks, and triggers alarms when necessary. Anomalies can include degradation of performance, crashes of components, errors during execution, toll fraud and so on.

Another application of ML related to system logs is the root cause determination problem. Using a training set comprised of error messages, different types of information messages and quality statistics published by the end users, the system can be trained to recognize the source of the problems depending on the recent log messages.



Based on the physical location and the load status of the media servers and end-user clients, ML can suggest the best media servers to place a given call or conference as well as predict the quality level, which can be provided by each topology.

Application level

Based on the sequence of presence status changes (free, on a call, in a meeting), ML can predict with some probability the status in the future. For example, it could inform you that the meeting you had planned for 9 a.m. on Monday with participant X, who is normally busy with a regular call at that time, even if this call is not on his calendar.

ML considers the history of routing and real-time state of the system and pairs customers with those contact center agents that most probably will give the best result. Customer transactional history includes preferred communication channel, past product purchases and service requests. This information is combined with agent profiles, skills and presence to predict the best customer-agent match.

Conversation content

Having text from chats, emails, files, or as a transcription of the calls and conferences, ML processes the content of these interactions and accelerates productivity by automatically capturing meeting notes from any to-do items, follow-ups, and important moments. If necessary, the transcription and meeting notes can be translated to different languages using real-time translation ML systems. For better quality of service, ML needs to use corporate domain-specific knowledge bases, like FAQs, in order to understand domain specific terms.

In the future, ML will be able to function proactively. Instead of responding to an explicit request to a bot's request, "provide me with the list of flights from Paris to New York," an ML-enabled system will observe the content of the discussion and show the list when it's needed by the conversation's flow.

Sentiment analysis is one of the most popular applications of ML in the scope of customer service organizations. It provides the information of what emotions were revealed by the customers while they were typing the messages.

AI ruled and AI assisted communications

AI supports conversations in automatic or semi-automatic mode.

Automatic mode is used in chat bots. A chat bot is a component of the system which is able to conduct a conversation by means of text or voice. Chat bots in enterprise context are typically used for different practical purposes like access to a knowledge base or access to some service. Chat bots differ by their level of intelligence. The most primitive ones don't engage AI, they just provide several predefined options from which a user can select. More advanced bots scan for keywords in the input and base their logic on found ones, often functioning like a search engine. The most complicated bots are full-fledged NLP (Natural Language Processing) engines, capable of understanding the user intent and acts accordingly.

One of the main use cases for chat bots is a representative function. That is, a chat bot is placed on the company website, where it waits for the clients and their requests. Such chat bots provide customers with access to company information as well as to company services. This approach allows businesses to optimize the number of employees in contact centers and customer service departments.

Another possible use is internal enterprise bots, providing access to some corporate functions like HR or IT support. For the moment, it's a less popular approach, but it's gaining traction, as it permits optimization of internal business processes. Voice bots, that are also known as virtual assistants, can be used to send commands to the system. This functionality is used, for example, in conferencing solutions, in which a user can speaks naturally to a conference solution to initiate, join or finish a meeting.

Semi-automatic AI supported conversations were first developed for contact center agents interacting with the customers. Applying AI techniques to the context of the interaction, the semi-automatic system provides an agent with the most pertinent answers for the given conversation, but the final decision remains for the human agent. A semi-automatic approach enables reusing of existing information and eliminates needing to do the same thing twice, and at the same time, involvement of a real agent guarantees the high level of meaningfulness of the interaction.

General company context

General company context is a set of information that is not a part of an enterprise communications system and hence, must be accessed in other enterprise systems to be used for communications related-functionality. For example:

- Employee information from an HR database
- List of projects, their description and resources dedicated to them
- Content of company documents, knowledge base, FAQ, ERP, etc.
- Physical location of a user, defined by his/her mobile device

ML can compile this information to optimize project execution by revealing to the participants that:

- Questions they pose are answered in a specific document
- Skills they searched for are available from a given person
- Customers and partners who were engaged in a previous project

Conclusion

While often perceived as a futuristic concept, artificial intelligence and machine learning are optimizing and improving an increasing number of business processes in a modern enterprise. Those, which are not so modern, risk losing competitiveness. There is an innovative level of quality and functionality thanks to AI. Customers, which will soon be acquainted with these new levels of the service, will refuse to use traditional ones.

The data, generated by a modern enterprise communications solution is vast and varied. From networking packets to discussion content, from application logs to recognized gestures of video conference participants – this contributes to the overall semantically rich depiction of user intentions and actions.

Applying AI to this data gives a clear advantage to an enterprise in terms of features and quality.

Alcatel·Lucent Enterprise