GAMAKA
Artificial Intelligence Solutions

DATA SCIENCE

INTERVIEW PREPARATION

DAY 20

www.gamakaai.com
+91-7378483656 +91-7378493293

Q1. Do you have any idea about Event2Mind in NLP?

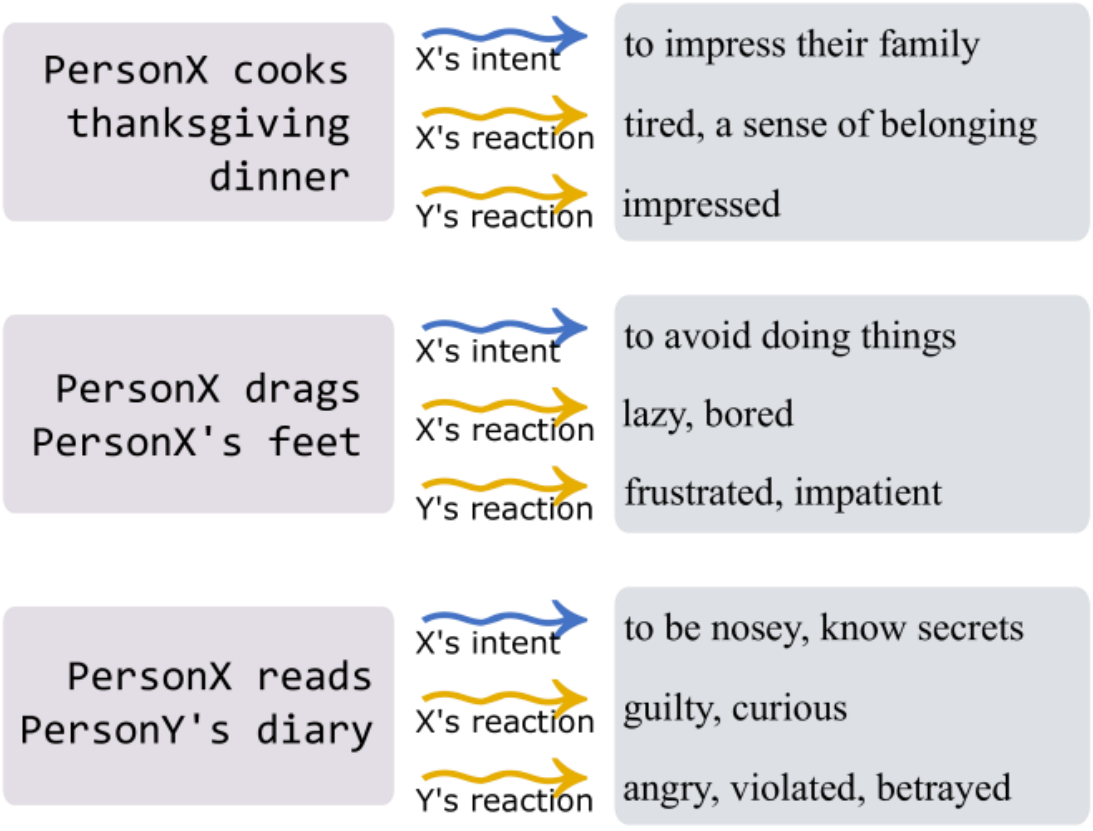
Answer:

Yes, it is based on NLP research paper to understand the common-sense inference from sentences.

Event2Mind: Common-sense Inference on Events, Intents, and Reactions

The study of "Commonsense Reasoning" in NLP deals with teaching computers how to gain and employ common sense knowledge. NLP systems require common sense to adapt quickly and understand humans as we talk to each other in a natural environment.

This paper proposes a new task to teach systems commonsense reasoning: given an event described in a short "event phrase" (e.g. "PersonX drinks coffee in the morning"), the researchers teach a system to reason about the likely intents ("PersonX wants to stay awake") and reactions ("PersonX feels alert") of the event's participants.



www.gamakaai.com
+91-7378483656 +91-7378493293

Understanding a narrative requires common-sense reasoning about the mental states of people in relation to events. For example, if “Robert is dragging his feet at work,” pragmatic implications about Robert’s *intent* are that “Robert wants to avoid doing things” (Above Fig). You can also infer that Robert’s *emotional reaction* might be feeling “bored” or “lazy.” Furthermore, while not explicitly mentioned, you can assume that people other than Robert are affected by the situation, and these people are likely to feel “impatient” or “frustrated.”

This type of pragmatic inference can likely be useful for a wide range of NLP applications that require accurate anticipation of people’s intents and emotional reactions, even when they are not expressly mentioned. For example, an ideal dialogue system should react in empathetic ways by reasoning about the human user’s mental state based on the events the user has experienced, without the user explicitly stating how they are feeling. Furthermore, advertisement systems on social media should be able to reason about the emotional reactions of people after events such as mass shootings and remove ads for guns, which might increase social distress. Also, the pragmatic inference is a necessary step toward automatic narrative understanding and generation. However, this type of commonsense social reasoning goes far beyond the widely studied entailment tasks and thus falls outside the scope of existing benchmarks.

Q2. What is SWAG in NLP?

Answer:

SWAG stands for **S**ituations **w**ith **A**dversarial **G**enerations is a dataset consisting of 113k multiplechoice questions about a rich spectrum of grounded situations.

Swag: A Large Scale Adversarial Dataset for Grounded Commonsense Inference

According to NLP research paper on SWAG is “Given a partial description like “he opened the hood of the car,” humans can reason about the situation and anticipate what might come next (“then, he examined the engine”). In this paper, you introduce the task of grounded commonsense inference, unifying natural language inference(NLI), and common-sense reasoning.

www.gamakaai.com
+91-7378483656 +91-7378493293

We present SWAG, a dataset with 113k multiple-choice questions about the rich spectrum of grounded positions. To address recurring challenges of annotation artifacts and human biases found in many existing datasets, we propose AF(Adversarial Filtering), a novel procedure that constructs a de-biased dataset by iteratively training an ensemble of stylistic classifiers, and using them to filter the data. To account for the aggressive adversarial filtering, we use state-of-the-art language models to oversample a diverse set of potential counterfactuals massively. Empirical results present that while humans can solve the resulting inference problems with high accuracy (88%), various competitive models make an effort on our task. We provide a comprehensive analysis that indicates significant opportunities for future research.

When we read a tale, we bring to it a large body of implied knowledge about the physical world. For instance, given the context “on stage, a man takes a seat at the piano,” we can easily infer what the situation might look like: a man is giving a piano performance, with a crowd watching him. We can furthermore infer him likely next action: he will most likely set his fingers on the piano key and start playing.

This type of natural language inference(NLI) requires common-sense reasoning, substantially broadening the scope of prior work that focused primarily on linguistic entailment. Whereas the dominant entailment paradigm asks if 2 natural language sentences (the ‘premise’ and the ‘hypothesis’) describe the same set of possible worlds, here we focus on whether a (multiple-choice) ending represents a possible (*future*) world that can a from the situation described in the premise, even when it is not strictly entailed. Making such inference necessitates a rich understanding of everyday physical conditions, including object affordances and frame semantics.

On stage, a woman takes a seat at the piano. She

a) sits on a bench as her sister plays with the doll.
b) smiles with someone as the music plays.
c) is in the crowd, watching the dancers.
d) nervously sets her fingers on the keys.

A girl is going across a set of monkey bars. She

a) jumps up across the monkey bars.
b) struggles onto the monkey bars to grab her head.
c) gets to the end and stands on a wooden plank.
d) jumps up and does a back flip.

The woman is now blow drying the dog. The dog

a) is placed in the kennel next to a woman's feet.
b) washes her face with the shampoo.
c) walks into frame and walks towards the dog.
d) tried to cut her face, so she is trying to do something very close to her face.

Table 1: Examples from Swag; the correct answer is **bolded**. Adversarial Filtering ensures that stylistic models find all options equally appealing.

Q3. What is the Pix2Pix network?

Answer:

Pix2Pix network: It is a Conditional GANs (cGAN) that learn the mapping from an input image to output an image.

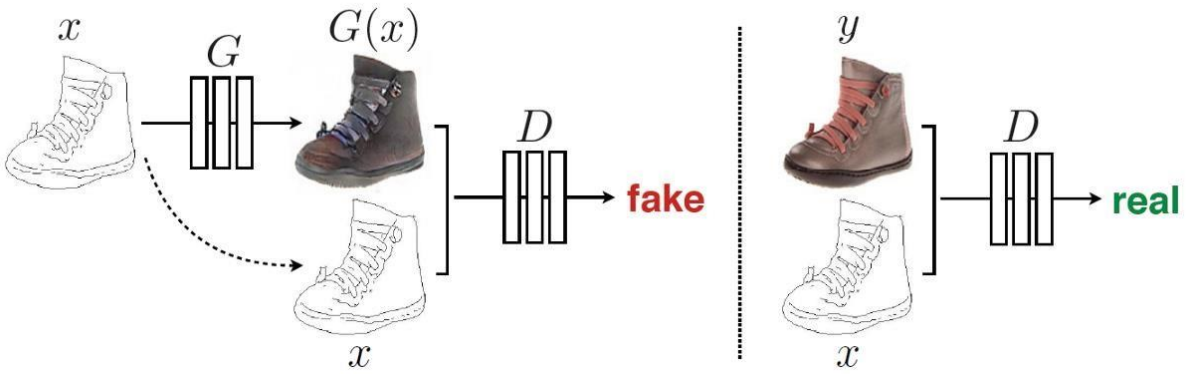
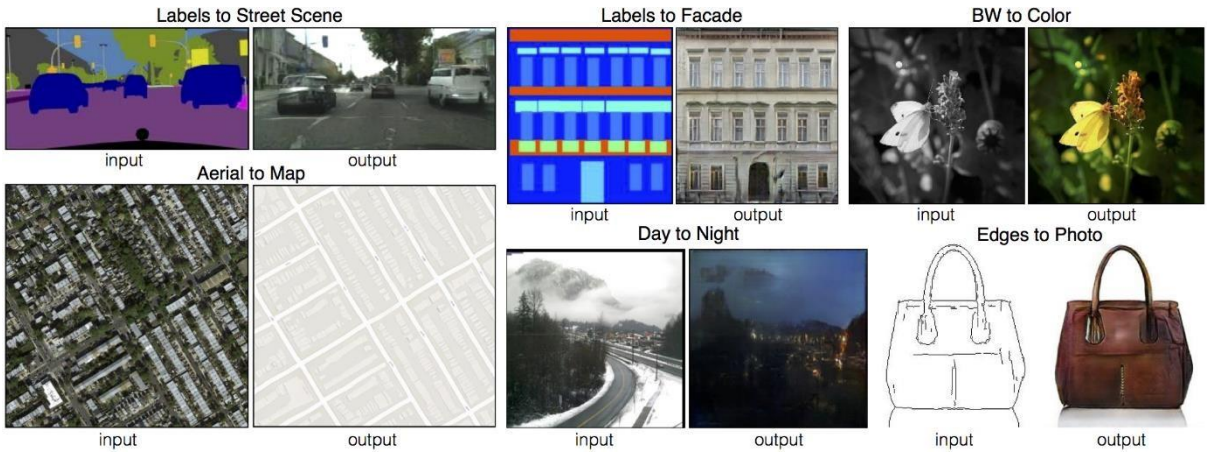


Image-To-Image Translation is the process for translating one representation of the image into another representation.

The image-to-image translation is another example of a task that GANs (Generative Adversarial Networks) are ideally suited for. These are tasks in which it is nearly impossible to hard-code a loss function. Studies on GANs are concerned with novel image synthesis, translating from a random vector z into an image. Image-to-Image translation converts one image to another like the edges of the bag below to the photo image. Another exciting example of this is shown below:



In Pix2Pix Dual Objective Function with an Adversarial and L1 Loss

www.gamakaai.com
+91-7378483656 +91-7378493293

A naive way to do Image-to-Image translation would be to discard the adversarial framework altogether. A source image would just be passed through a parametric function, and the difference in the resulting image and the ground truth output would be used to update the weights of the network. However, designing this loss function with standard distance measures such as L1 and L2 will fail to capture many of the essential distinctive characteristics between these images. However, authors do find some value to the L1 loss function as a weighted sidekick to the adversarial loss function.

The Conditional-Adversarial Loss (Generator versus Discriminator) is very popularly formatted as follows:

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x,z))],$$

The L1 loss function previously mentioned is shown below:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1].$$

Combining these functions results in:

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_{L1}(G).$$

In the experiments, the authors report that they found the most success with the lambda parameter equal to 100.

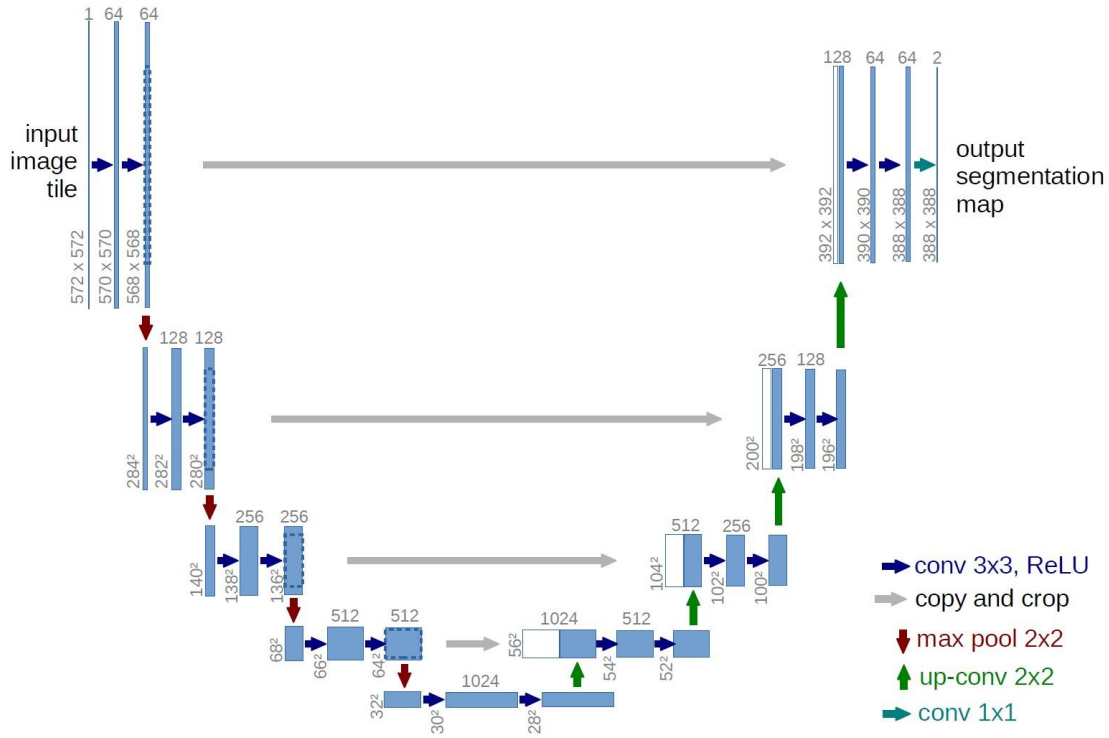
www.gamakaai.com
+91-7378483656 +91-7378493293

Q4. Explain UNet Architecture?

Answer:

U-Net architecture: It is built upon the Fully Convolutional Network and modified in a way that it yields better segmentation in medical imaging. Compared to FCN-8, the two main differences are (a) U-net is symmetric and (b) the skip connections between the downsampling path and upsampling path apply a concatenation operator instead of a sum. These skip connections intend to provide local information to the global information while upsampling. Because of its symmetry, the network has a large number of feature maps in the upsampling path, which allows transferring information. By comparison, the underlying FCN architecture only had the *number of classes* feature maps in its upsampling way.

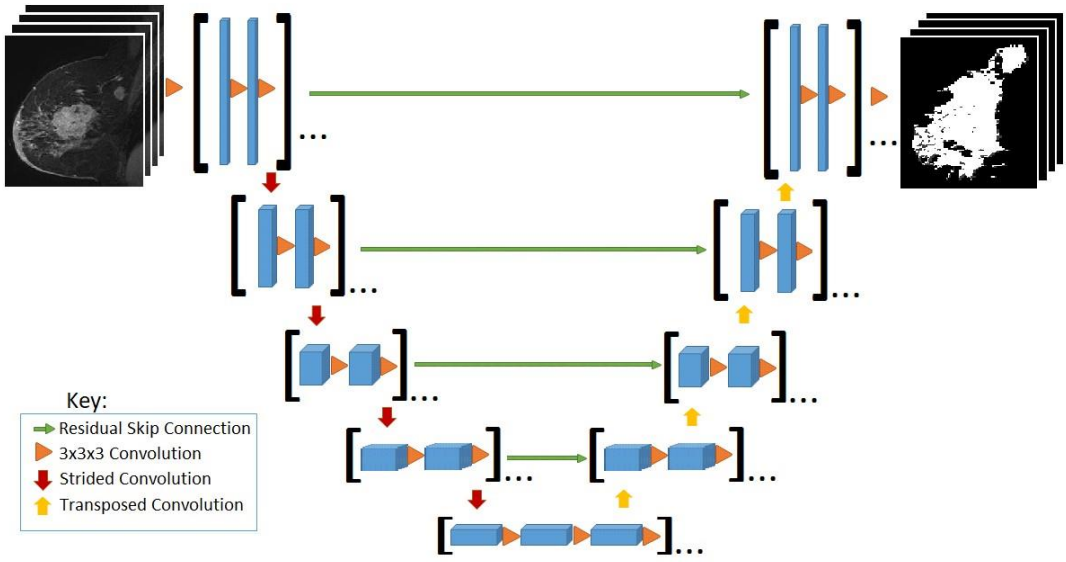
How does it work?



The UNet architecture looks like a 'U,' which justifies its name. This UNet architecture consists of 3 sections: The contraction, the bottleneck, and the expansion section. The contraction section is made of many contraction blocks. Each block takes an input that applies two 3X3 convolution layers, followed by a 2X2 max pooling. The number of features or kernel maps after each block

doubles so that UNet architecture can learn complex structures. Bottommost layer mediates between the contraction layer and the expansion layer. It uses two 3X3 CNN layers followed by 2X2 up convolution layer.

But the heart of this architecture lies in the expansion section. Similar to the contraction layer, it also has several expansion blocks. Each block passes input to two 3X3 CNN layers, followed by a 2X2 upsampling layer. After each block number of feature maps used by the convolutional layer, get half to maintain symmetry. However, every time input is also get appended by feature maps of the corresponding contraction layer. This action would ensure that features that are learned while contracting the image will be used to reconstruct it. The number of expansion blocks is as same as the number of contraction blocks. After that, the resultant mapping passes through another 3X3 CNN layer, with the number of feature maps equal to the number of segments desired.



www.gamakaai.com
+91-7378483656 +91-7378493293

Q5. What is pair2vec?

Answer:

This paper pre trains *word pair representations* by maximizing pointwise mutual information of pairs of words with their context. This encourages a model to learn more meaningful representations of word pairs than with more general objectives, like modeling. The pre-trained representations are useful in tasks like SQuAD and MultiNLI that require cross-sentence inference. You can expect to see more pretraining tasks that capture properties particularly suited to specific downstream tasks and are complementary to more general-purpose tasks like language modeling.

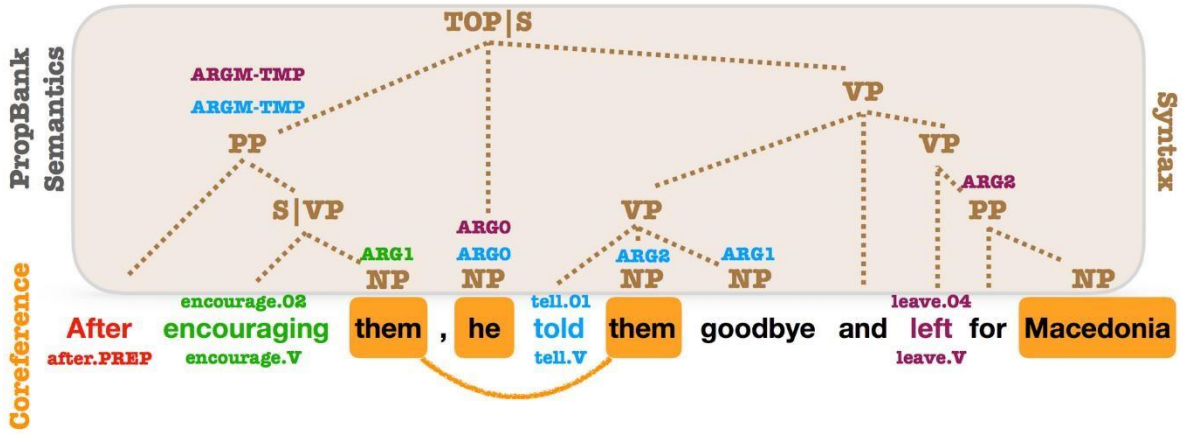
Reasoning about implied relationships between pairs of words is crucial for cross sentences inference problems like question answering (QA) and natural language inference (NLI). In NLI, e.g., given a premise such as "*golf is prohibitively expensive*," inferring that the hypothesis "*golf is a cheap pastime*" is a contradiction requires one to know that *expensive* and *cheap* are antonyms. Recent work has shown that current models, which rely heavily on unsupervised single-word embeddings, struggle to grasp such relationships. In this pair2vec paper, we show that they can be learned with word pair2vec(pair vector), which are trained, unsupervised, at a huge scale, and which significantly improve performance when added to existing cross-sentence attention mechanisms.

X	Y	Contexts
		with X and Y baths
hot	*cold*	too X or too Y
		neither X nor Y
		in X, Y
Portland	*Oregon*	the X metropolitan area in Y
		X International Airport in Y
		food X are maize, Y, etc
crop	*wheat*	dry X, such as Y,
		more X circles appeared in Y fields
		X OS comes with Y play
Android	*Google*	the X team at Y
		X is developed by Y

Table 1: Example word pairs (italicized) and their contexts (Wikipedia).

Unlike single word representations, which are typically trained by modeling the co-occurrence of a target word x with its context c, our word-pair

representations are learned by modeling the three-way co-occurrence between two words (x,y) and the context c that ties them together, as illustrated in above Table. While similar training signal has been used to learn models for ontology construction and knowledge base completion, this paper shows, for the first time, that considerable scale learning of pairwise embeddings can be used to improve the performance of neural crosssentence inference models directly.



Q6. What is Meta-Learning?

Answer:

Meta-learning: It is an exciting area of research that tackles the problem of learning to learn. The goal is to design models that can learn new skills or fastly to adapt to new environments with minimum training examples. Not only does this dramatically speed up and improve the design of ML(Machine Learning) pipelines or neural architectures, but it also allows us to replace hand-engineered algorithms with novel approaches learned in a data-driven way.

The goal of meta-learning is to train the model on a variety of learning tasks, such that it can solve new learning tasks with only a small number of training samples. It tends to focus on finding **model agnostic** solutions, whereas multi-task learning remains deeply tied to model architecture.

Thus, meta-level AI algorithms make AI systems:

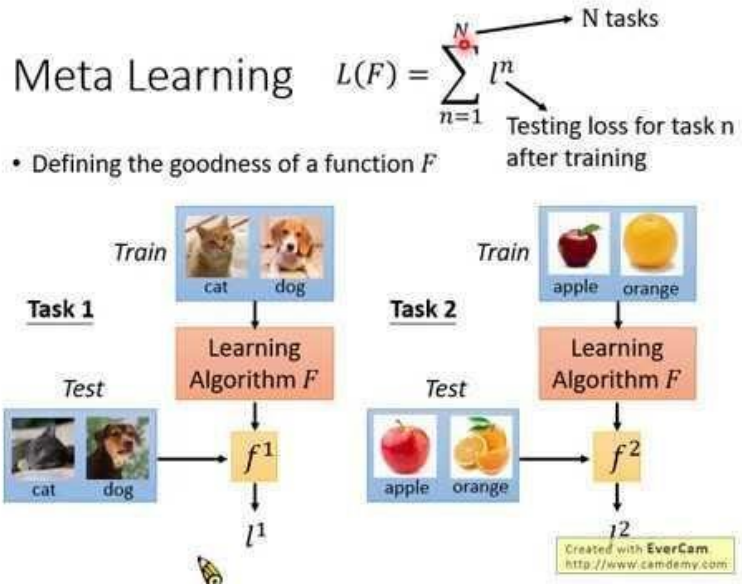
· Learn faster

· Generalizable to many tasks

· Adaptable to environmental changes like in Reinforcement Learning

www.gamakaai.com
+91-7378483656 +91-7378493293

One can solve any problem with a single model, but meta-learning should not be confused with oneshot learning.



Q7. What is ALiPy(Active Learning in Python)?

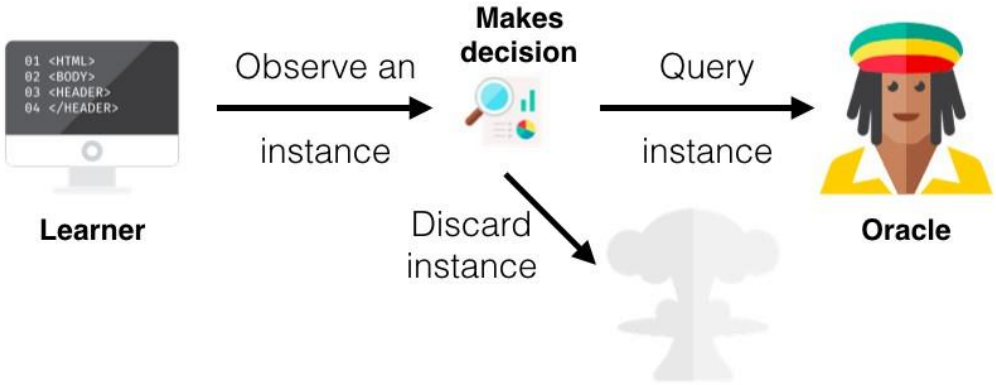
Answer:

Supervised ML methods usually require a large set of labeled examples for model training. However, in many real applications, there are ample unlabeled data but limited labeled data; and acquisition of labels is costly. Active learning (AL) reduces labeling costs by iteratively selecting the most valuable data to query their labels from the annotator.

Active learning is the leading approach to learning with limited labeled data. It tries to reduce human efforts on data annotation by actively querying the most prominent examples.

ALiPy is a Python toolbox for active learning(AL), which is suitable for various users. On the one hand, the entire process of active learning has been well implemented. Users can efficiently perform experiments by many lines of codes to finish the entire process from data pre-processes to

result in visualization. More than 20 commonly used active learning(AL) methods have been implemented in the toolbox, providing users many choices.



Q8.What is the Lingvo model?

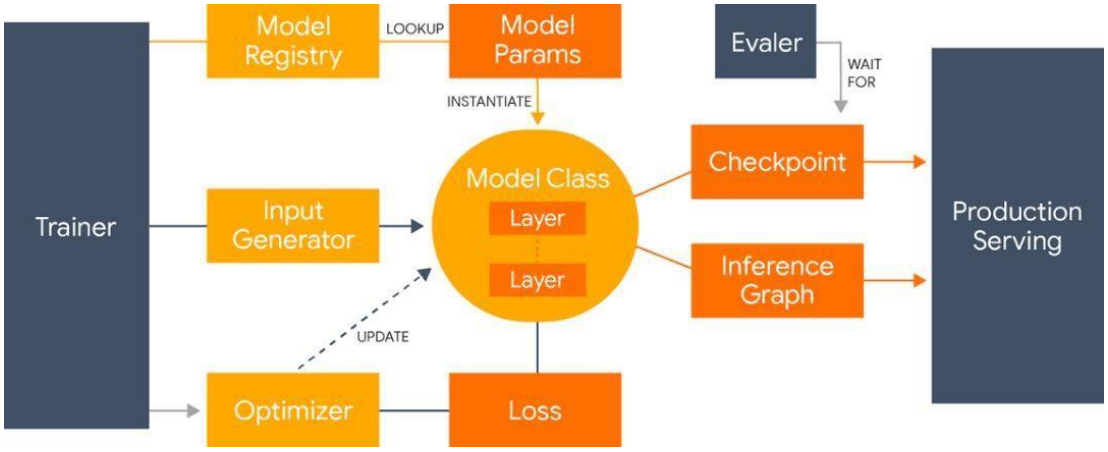
Answer:

Lingvo: It is a Tensorflow framework offering a complete solution for collaborative deep learning research, with a particular focus towards sequence-to-sequence models. These models are composed of modular building blocks that are flexible and easily extensible, and experiment configurations are centralized and highly customizable. Distributed training and quantized inference are supported directly within a framework, and it contains existing implementations of an ample number of utilities, helper functions, and newest research ideas. This model has been used in collaboration by dozens of researchers in more than 20 papers over the last two years.

Why does this Lingvo research matter?

The process of establishing a new deep learning(DL) system is quite complicated. It involves exploring an ample space of design choices involving training data, data processing logic, the size, and type of model components, the optimization procedures, and the path to deployment. This complexity requires the framework that quickly facilitates the production of new combinations and the modifications from existing documents and experiments and shares these new results. It is a workspace ready to be used by deep learning researchers or developers. Nguyen Says: “We have researchers working on state-of-the-art(SOTA) products and research algorithms, basing their research off of the same codebase. This ensures that code is battle-tested.

Our collective experience is encoded in means of good defaults and primitives that we have found useful over these tasks."



Q9. What is Dropout Neural Networks?

Answer:

The term "dropout" refers to dropping out units (both hidden and visible) in a neural network.

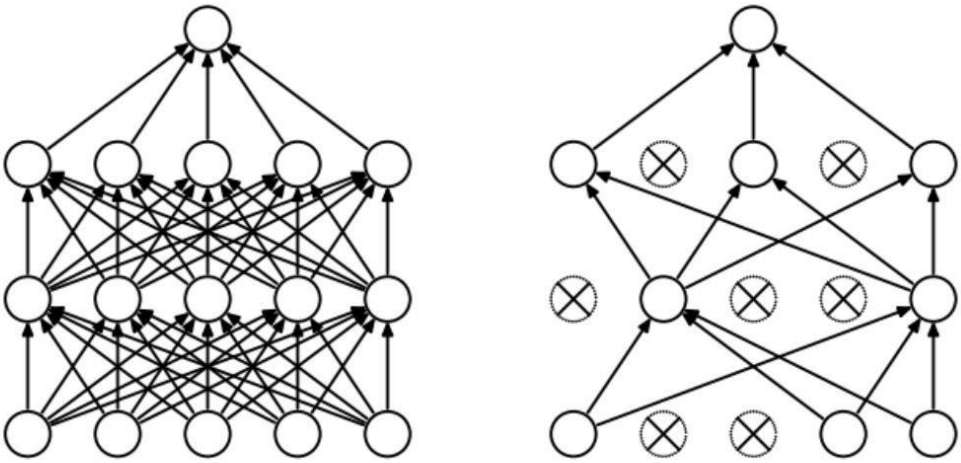
At each training stage, individual nodes are either dropped out of the net with probability 1-p or kept with probability p, so that a reduced network is left; incoming and outgoing edges to a dropped-out node are also removed.

Why do we need Dropout?

The answer to these questions is "to prevent over-fitting."

A fully connected layer occupies most of the parameters, and hence, neurons develop co-dependency amongst each other during training, which curbs the individual power of each neuron leading to overfitting of training data.

www.gamakaai.com
+91-7378483656 +91-7378493293

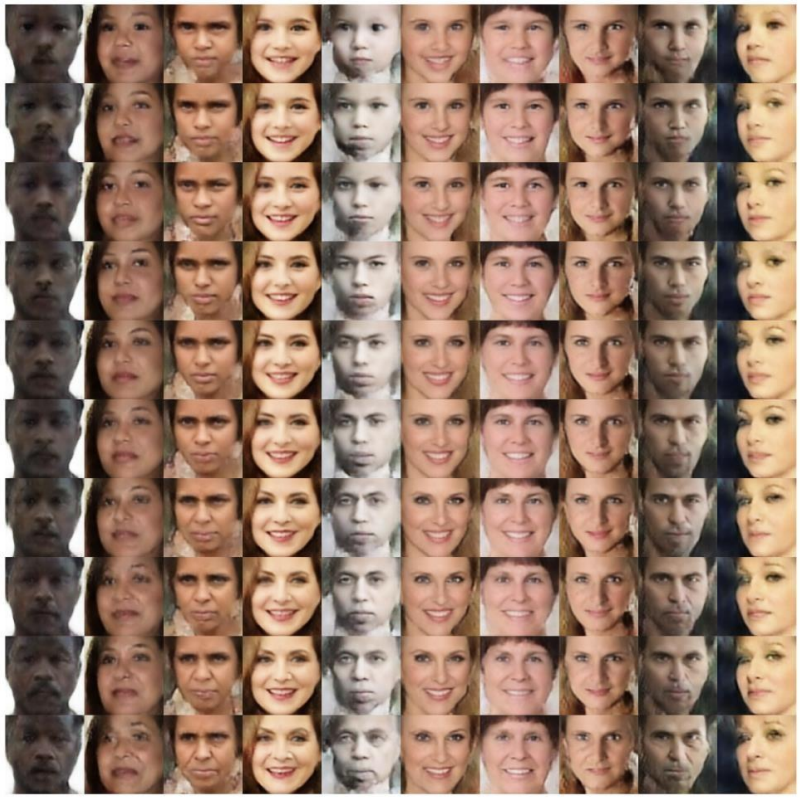


(a) Standard Neural Net (b) After applying dropout.

Q10. What is GAN?

Answer:

A **generative adversarial network** (**GAN**): It is a class of machine learning systems invented by Ian Goodfellow and his colleagues in 2014. Two neural networks are contesting with each other in a game (in the idea of game theory, often but not always in the form of a zero-sum game). Given a training set, this technique learns to generate new data with the same statistics as the training set. E.g., a GAN trained on photographs can produce original pictures that look at least superficially authentic to human observers, having many realistic characteristics. Though initially proposed as a form of a generative model for unsupervised learning, GANs have also proven useful for semisupervised learning,[2] fully supervised learning, and reinforcement learning. Example of GAN



- Given an image of a face, the network can construct an image that represents how that person could look when they are old.

Generative Adversarial Networks takes up a game-theoretic approach, unlike a conventional neural network. The network learns to generate from a training distribution through a 2-player game. The two entities are Generator and Discriminator. These two adversaries are in constant battle throughout the training process.

www.gamakaai.com
+91-7378483656 +91-7378493293