2. Leia o artigo Review of deep learning: concepts, CNN architectures, challenges, applications, future directions (Alzubaidi, L. 2021). Baseado no artigo, descreva os principais desafios e limitações de DL. Quais alternativas os autores dão aos problemas?

Staying current with the media's coverage of the release of the latstes version of ChatGPT it may seem like AI (Artificial Intelligence) is a brand new technology. This is obviously not the case and subsets of the AI family, such as ML (machine learning) have been used by industries such as healthcare, or in research for things like environmental science for years. Even then some of the concepts can date back decades and with modern hardware and advancements of research they are finally becoming applicable enough for implementation.

DL (Deep Learning) is also a concept that has been around for a while. However, recently it has shown that it can be used to create models or provide classification that out performs ML with less upfront cost or work for it to be viable for implementation. With the adoption of DL there have been limitations and challenges that have arrived before it became the standard and after it has been implemented. These issues are well detailed in the article "Review of deep learning: concepts, CNN architectures, challenger, applications, future direction" by Alzubaidi et al.

These challenges involve or are labeled by training data, imbalanced data, interpretability of data, uncertainty scaling, catastrophic forgetting, model compression, overfitting, vanishing gradient problem, exploding gradient problem, and underspecification.

DL implementation requires vastly more data than ML --- which of course can be an issue for new applications where the data doesn't exist yet or for niche industries where there isn't a pool of readily available data ready for the model. This can be addressed through various means including the transfer-learning concept, data augmentation or simulated data.

Transfer-learning is the use of another well-trained model and changing two or even one while considering the current or new application. Another technique commonly used is data Augmentation. This can involve things like using flipping, color space, cropping, rotation, translation and noise injection. These need to be used with care depending on the application as they may change the data in way that makes it incorrect or illegible thus reducing the performance of the model. Lastly creating simulated data is another way to increase the amount of data in your training set.

Imbalanced data is another problem faced and is commonly found in applications where a positive or negative result is more frequently found. Thus, making the training set to have bias towards one of the classes. This can be addressed by down sampling the larger class or using other techniques to sample up the smaller class.

Interpretability of data can be an issue for machine learning due to the specialization of the classification. The paper gives an example of disease diagnosis because you would need highly trained people to interpret the data to ensure the model is trained on the correct information. Some solutions include back-propagation or perturbation-based approaches. These can be computationally intensive, however, they are fairly simple to understand.

Uncertainty scaling is another issue of needing to include the uncertainty for measurements. Uncertainty scaling is a technique used in Bayesian Deep Learning to improve the calibration of uncertainty estimates produced by the model. In Bayesian Deep Learning, uncertainty refers to the model's confidence or certainty about its predictions. Accurate uncertainty estimates are crucial in various applications, such as decision-making in safety-critical systems and in tasks with limited labeled data.

In traditional deep learning, uncertainty is typically not explicitly modeled, and the model outputs deterministic predictions. Bayesian Deep Learning, on the other hand, treats model weights as random variables and captures uncertainty in predictions by considering multiple possible outcomes.

Uncertainty scaling aims to improve the reliability of uncertainty estimates by scaling or calibrating them based on observed confidence levels. The main idea is to align the predicted uncertainties with the actual performance of the model.

Catastrophic forgetting can be an issue when a model is trained to classify X number of things. The paper provides an example of plant species identification. It is possible that when one more class is added that the model will then perform very poorly after this. Of course, one solution is to retrain the model however that is computationally and labor intensive. There are ML techniques that will help this, and they include regularizations such as EWC, rehearsal training techniques and dynamic neural network architecture like iCaRL and lastly dual-memory learning systems.

Model compression is another problem mentioned in the article, which is faced when computational and memory requirements may be too high for certain environments. This can happen in any industry with models that require a large number of parameters but then need to run on simple hardware, which can cause a huge gap in desired performance. Redundant parameters, which are also called parameter pruning is a technique which can address this without significate performance impact. Similar to transfer learning, you can use a larger model that might be computationally intensive to write a smaller model. This technique is called knowledge distillation. It is also possible to reduce the number of parameters by using compact convolution filters.

Overfitting is a problem with any sort of statistical model, and this can happen in the training stage of the DL model. Weight decay is commonly used as a solution for a universal regularizer. Also, you can explore the inputs of the model to analyze data corruption and data augmentation as these may produce overfitting. Lastly exploring the output of the model, techniques can be used to penalize outputs that are over-confident.

Underspecification is another challenge that occurs when a DL model is deployed in its desired application and surprisingly it has severely degraded performance. Essentially the solution to this problem is to create stress tests before deployment that will go beyond a standard training set. This is very important for applications that are extremely complex and may have serious impacts on safety.

Lastly, one of the issues that arises when running backpropagation and gradient-based learning techniques, especially in large datasets is that, in every training, the weight updating might not occur due to a very small gradient. This could ultimately lead to the neural network shutting down. The number of layers used could then play an issue.

When gradients (derivatives of the loss function with respect to the model parameters) become extremely small as they are propagated backward through the network during the process of backpropagation, it presents a struggle for the model to update the weights effectively, leading to slow or stalled learning and poor convergence during training, especially when using activation functions that squash their input, such as the sigmoid or tanh functions.

One solution to the vanishing gradient problem is to use activation functions that alleviate the vanishing gradient issue. ReLU activation, for instance sets all negative values to zero and passes non-negative values as is, which allows for efficient learning and avoids the vanishing gradient problem for positive inputs. By using ReLU and its variants, the gradient remains non-zero for positive inputs, enabling effective weight updates and mitigating the vanishing gradient problem.

Another powerful solution is the use of skip connections, also known as residual connections. Skip connections allow gradients to bypass some layers and directly propagate to deeper layers during training. This mechanism helps in avoiding excessive multiplication of gradients through numerous layers, preventing the gradients from vanishing.

One popular architecture that uses skip connections is the ResNet (Residual Network). ResNet introduces shortcut connections that add the original input to the output of a residual block. By doing so, it facilitates the flow of gradients, making it easier for the network to learn deeper representations.

By using activation functions like ReLU and skip connections, deep neural networks can effectively address the vanishing gradient problem, allowing for more stable and efficient training, and enabling the successful training of very deep models.

Batch Normalization is another powerful technique used to address the vanishing gradient problem and improve the training of deep neural networks. The main idea behind Batch normalization is to normalize the input of each layer within a mini-batch during training. This normalization helps in mitigating the vanishing gradient problem and allows the network to train more effectively.

In a mini-batch of training samples, Batch Normalization computes the mean and standard deviation for each feature separately. It then normalizes the feature activations such that they have a mean close to zero and a standard deviation close to one.

The normalization process helps in preventing extreme values in the activations, which in turn prevents the gradients from becoming too small during backpropagation. This enables more stable and faster training of deep models, leading to improved convergence and better generalization.

Overall, all those techniques mentioned have been instrumental in achieving impressive results in various Deep Learning tasks.

3. Cite 3 aplicações de algoritmos de Deep Learning. Uma delas tem que usar uma solução com CNN e uma delas tem que usar uma solução de RNNs. Explique a aplicação e os resultados obtidos. Não deixe de citar a fonte utilizada, sendo preferencialmente artigos.

Application with CNN: Image Classification

Convolutional Neural Networks (CNNs) have become a dominant choice for image classification tasks due to their ability to automatically learn hierarchical features from images. A remarkable milestone for CNNs was the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), a prestigious competition in computer vision. During the ILSVRC, CNNs, particularly the ResNet (Residual Network) architecture, demonstrated remarkable success in classifying images into predefined categories.

Reference:

"Deep Residual Learning for Image Recognition" by Kaiming He et al. (2016).

Summary:

The ResNet architecture achieved groundbreaking results in the ILSVRC, outperforming traditional computer vision techniques. By employing residual connections, ResNet significantly reduced the classification error and effectively tackled the vanishing gradient problem. This breakthrough advanced the state-of-the-art in image classification, enabling more accurate and robust image recognition systems.

Application with RNNs: Natural Language Processing (NLP) - Language Translation

Recurrent Neural Networks (RNNs) have proven their effectiveness in processing sequential data, making them invaluable for Natural Language Processing tasks. One popular application of RNNs is language translation, where the Long Short-Term Memory (LSTM) network, a variant of RNNs, has gained prominence. In language translation tasks, the LSTM model takes a sequence of words from one language as input and generates the corresponding translated sequence in another language.

Reference:

"Sequence to Sequence Learning with Neural Networks" by Ilya Sutskever et al. (2014).

Summary:

LSTM-based sequence-to-sequence models have achieved promising results in machine translation tasks. Unlike traditional statistical machine translation approaches, LSTM models can handle variable-length input sequences and capture long-range dependencies effectively. They have exhibited near-human-level performance on various language pairs, allowing for more accurate and fluent translations across different languages.

Application with Deep Learning: Speech Recognition

Deep Learning models, encompassing both RNNs and CNNs, have revolutionized automatic speech recognition systems. These models process audio signals as input and convert them into text, making them crucial for speech-to-text applications. A notable architecture in speech recognition is the Listen, Attend, and Spell (LAS) model, which combines attention mechanisms with RNNs to effectively transcribe spoken language into written text.

Reference:

"Listen, Attend and Spell" by Chan et al. (2016).

Summary:

The LAS model has achieved state-of-the-art results in automatic speech recognition tasks, surpassing traditional methods. By employing attention mechanisms, the LAS model can focus on relevant acoustic features and phonetic patterns, allowing for more accurate and context-aware transcriptions. This advancement has significantly improved speech recognition systems, making them more adaptable to diverse accents, background noise, and challenging acoustic environments.