

Abstract:

Deep learning (DL) permite que modelos computacionais compostos de múltiplas camadas de processamento possam aprender representações de dados com múltiplas camadas de abstração. São citados as principais áreas em que deep learning se destacou, como reconhecimento e detecção de objetos em imagens e em reconhecimento de fala. O procedimento usado para o ajuste dos parâmetros internos baseia-se no algoritmo de retropropagação do erro, dado um conjunto de dados suficientemente grande. Para processamento de fala redes convolucionais profundas são usadas, enquanto que redes recorrentes são usadas para dados sequenciais, como texto e fala.

Introdução:

Por décadas foi necessário engenharia para se projetar extratores de features que transformassem dados crus para que uma técnica de machine learning pudesse ser usada. Em contraste, hoje consegue-se uma boa representação dos dados (em múltiplas camadas) com deep learning, partindo apenas dos mesmos dados crus. Por fim, os autores acreditam que o sucesso maior em deep learning ainda está por vir, pelo motivo de exigir pouco ou nada de trabalho manual, bem como a crescente quantidade de dados e poder computacional.

Aprendizado supervisionado:

Aqui é explicado como acontece o aprendizado supervisionado, partindo de um dataset rotulado, escolha da função de custo e ajuste dos pesos utilizando o vetor gradiente. Também é mostrado o conceito de mini-batch, onde usa-se o gradiente estocástico descendente (SGD), calculado a partir de alguns padrões, para se fazer um ajuste dos pesos. É apresentado como um classificador linear (ou qualquer outro “raso”) funciona, e o porquê deles necessitarem de um bom extrator de features, baseando-se no dilema seletividade-invariância. Para se projetar um extrator de features, se faz necessário domínio e habilidades, enquanto que com o uso de deep learning, features são aprendidas automaticamente e de forma a aumentar tanto seletividade quanto a invariância da representação.

Retropropagação para treinar arquiteturas multicamadas:

Nesta seção ilustra-se como é o processo de retropropagação do erro para ajuste dos pesos, dada uma função de perda. No fim da década de 90, as redes neurais foram praticamente esquecidas, voltando a tona em meados de 2006, quando foi introduzido aprendizado não-supervisionado para extração de features. Com o uso de GPUs para acelerar os treinamentos das redes, grandes resultados foram alcançados em reconhecimento de fala.

Redes neurais convolucionais:

As ConvNets, são mais fáceis de treinar devido ao número reduzido de parâmetros e aproveitam a relação de vizinhança nos dados. As camadas iniciais são compostas de convoluções e pooling. A convolução permite a invariância de localização de sinais (ex: objeto em imagem), e as unidades de pooling suavizam pequenas distorções e deslocamentos.

Entendimento de imagens com redes convolucionais profundas:

Apesar do sucesso das ConvNets elas foram esquecidas, voltando em cena em 2012 com a rede AlexNet na competição ImageNet. Diversas técnicas foram usadas: uso de GPU, ReLUs, data augmentation, dropout. Empresas de software hoje empregam ConvNets em produtos e serviços e empresas de hardware desenvolvem chips próprios para este uso.

Representações distribuídas e processamento de linguagem:

A teoria por trás de DL mostra que aprender representações de forma distribuída, apresenta duas vantagens exponenciais com relação aos métodos que não usam representação distribuída: através das combinações das features e das camadas. No caso de predição da próxima palavra em PLN, na primeira camada cada palavra cria um padrão de ativação, nas seguintes, a rede aprende a converter os vetores de palavra de entrada em um vetor de palavra de saída para a próxima palavra predita. No fundo o que a rede está fazendo é aprender a extrair características semânticas de um texto, embora isso nunca foi dado como entrada, mas sim descoberto pela rede durante o processo de treinamento. Antes de DL em linguagem, os modelos não exploravam essas representações distribuídas. Eles eram baseados na frequência de ocorrência de sequências de símbolos (N-grams), que trata cada palavra separadamente. Portanto, não generalizam considerando a semântica, algo que os modelos de linguagem neural fazem ao levar as palavras para o espaço de vetores.

Redes neurais recorrentes:

Para dados sequenciais, como fala e escrita, as RNNs são comumente usadas. As unidades escondidas mantêm um vetor de estados, o que faz com que as RNN apresente dinâmica. Aqui é introduzida também a rede LSTM, que se mostraram mais efetivas que as RNN para determinados casos.

O futuro de deep learning:

Por fim os autores esperam que o aprendizado não-supervisionado se tornem mais importante a longo prazo. Também afirmar que PLN é uma área em que DL está fadada a ter grande impacto nos próximos anos.