

1. Introdução

Inicialmente, devemos começar descrevendo as diferenças entre os tipos de aprendizagem, para nos ajudar a entender a situação de motivação.

- Aprendizagem não supervisionada: trabalha apenas com dados não rotulados, buscando encontrar padrões, agrupamentos ou estruturas intrínsecas nos dados. Exemplo: é o agrupamento (clustering)
- Aprendizagem supervisionada: utiliza um conjunto de treinamento no qual cada exemplo está associado ao seu rótulo ou valor alvo. Exemplo: classificação de imagens.
- Aprendizagem semi-supervisionada: utiliza-se uma parcela de dados rotulados (pequena) e uma grande quantidade de dados não rotulados, a fim de potencializar o treinamento do modelo e melhorar a generalização.

Motivação: Os dados são abundantes, porém, rotular cada um deles, é uma tarefa muito trabalhosa. A aprendizagem semi-supervisionada entra como uma solução para esta problemática, pois ela reduz custos, explorando os dados não rotulados em conjunto com o pequeno conjunto rotulado.

2. Treinamento

2.1 - Self-Training

- O modelo inicial é treinado com base nos dados rotulados. Com isso, o modelo consegue atribuir pseudo-rótulos aos dados não rotulados com uma maior confiança.
- Os dados pseudo-rotulados, são adicionados ao conjunto de treinamento, com esse processo se repetindo.

2.2 - Co-Training

- Existem 2 visões independentes dos dados
- Dois classificadores são treinados separadamente; cada classificador rotula os exemplos com maior confiança para o outro.
- É útil quando existe complementaridade nas diferentes visões ou conjuntos de atributos.

3. Vantagens e limitações

3.1 - Vantagens

- Redução de custos, devido a menor dependência de rótulos.
- Melhor generalização, devido ao maior acesso à distribuição de dados.
- Variedade de aplicações, como visão computacional, linguagem natural.

3.2 - Limitações

- Propagação de erros no caso dos pseudo-rótulos serem imprecisos, propagando isso pro restante do modelo.
- Complexidade Computacional métodos avançados (por exemplo, baseados em grafos ou redes neurais profundas) demandam poder de processamento significativo

4. Aplicações

4.1 - Visão computacional

- Útil na classificação de imagens médicas, pois para rotulá-las mais assertivamente, precisa de um médico.
- O pequeno conjunto de imagens rotuladas pelos médicos, orienta o modelo na classificação de uma grande massa de dados não rotulados.

4.2 - Reconhecimento de voz

- Um modelo inicial é treinado em um pequeno conjunto de áudios transcritos; depois, as predições em áudios não rotulados são utilizadas para expandir o treinamento, aprimorando a acurácia do sistema de reconhecimento de fala.

5. Conclusão

A Aprendizagem Semi-Supervisionada é uma solução estratégica para problemas em que dados rotulados são escassos ou custosos, mas há grande disponibilidade de dados não rotulados.

À medida que o volume de dados não rotulados continua a crescer de forma exponencial, a Aprendizagem Semi-Supervisionada tende a se tornar cada vez mais relevante em diversas áreas.

