We’ve already trained several models with classic and novel approaches. But in a real-world sentiment analysis task a perfectly labelled training set cannot be guaranteed. We can break down this challenge into 2 situations: 1. Training set comes partially labelled or doesn’t have rating scores at all. 2. The training set is labelled with noisy annotations.

For the first scenario we have 3 ways to tackle the problem:

1. Co-Training (Blum & Mitchell)

If the training dataset comes partially labelled and samples have two views which mutually redundant (e.g. a dataset of people where samples represented by their ID and facial features). 2 classifiers can be trained on each view and later will be used to predict unlabeled subset to augment the labelled one. This method assumes there are 2 feature dimensions which mutually uncorrelated, so it is not suitable for our task but worth mentioning.

1. Transfer learning with Pre-Trained Models (Devlin et al.)

Pre-trained models like BERT, which usually haven been trained on a huge amount of data. This approach has been proven performs great on none to minimal labeled dataset.

1. Seed Labeling ((Hogenboom et al.))

Manually tag some samples in the dataset or utilize existing sentiment lexicons to identify sentimental words. Craft rules based on the frequency of sentiments-bearing words. Use these lexicons and rules to automatically create pseudo-labels.

If the training set is noisily annotated, we also have 3 approaches, namely:

1. Loss Function Correction (Patrini et al.))

Adjust the loss function during the training stage to account for noisy labels and thus minimize its effect on the final model. Assume there is a Noise Transition Matrix T representing the probability of a true label being transformed into another. This matrix can be estimated using a clean noise-free validation set. But usually, the estimation is not very effective and far from the genuine one.

1. Noise Adaptation Layer (Goldberger & Ben-Reuven)

Include a specialized layer within the model architecture which will learn how to adapt and correct for the label noise. This layer is usually a fully connected layer before the final output and will dynamically learn the noise distribution during the training stage. This mechanism is simple yet effective and tackle the problem internally within the model without explicit assumptions.

1. Ensemble Methods (Zhou)

Train multiple weaker models that parallelly make prediction and take the majority vote before final output.

**References**

1. Blum, A., & Mitchell, T. *Combining Labeled and Unlabeled Data with Co-Training.*
2. Devlin, J., et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.*
3. Hogenboom, A., et al. *A Lexicon-Based Approach to Sentiment Analysis for Twitter Data.*
4. Patrini, G., et al. *Making Deep Neural Networks Robust to Label Noise: A Loss Correction Approach.*
5. Goldberger, J., & Ben-Reuven, E. *Training Deep Neural-Networks Using a Noise Adaptation Layer.*
6. Zhou, Z.-H. *Ensemble Methods: Foundations and Algorithms.*