A Recurrent Neural Pipeline for Multi-Class | Multi-Label Text Classification

The focus of this tutorial is to understand the problem of Multiclass Multilabel (MM) Text Classification in NLP and talk about tangible concepts which can be leveraged using state-of-the-art tools and techniques to build deep learning models to tackle this problem. We will be giving a conceptual overview of what MM Text Classification entails followed by our approach around it. Once the audience gets some foundational knowledge around MM Text Classification, we will showcase our novel approach in handling such problem statements using Deep Learning via a Hands-on-Session.

Sample Problem: Identifying Categories & Sub-Categories from Text.
Example: Stack-overflow Question Classification into Categories (NLP/Computer Vision/Arts) and Sub-categories (embeddings/neural-network/model-weights/design)

Overall, the tutorial will be structured as follows:

Part 1: Understanding: Problem Statement, Dataset & ML approach (15 min)

- Framing a Multiclass Multilabel classification problem statement.
- Intuition of some use cases around it. What happens when classes are not mutually exclusive and how to tackle it?
- Building up a dataset from a small True Corpus. Data extrapolation for augmenting training data.
- Define the problem: Text classification vs Topic modelling.
- Key takeaway – Problem Framing.

Part 2: Capturing Relations in Text | Embeddings (15 min)

- What are Embedding? How do get document embedding from word embedding.
- Highlight SOTA embeddings and when to use them, e.g. BERT, ELMO, FLAIR, Glove, etc.
- Understand stacking of the embeddings to capture generic & custom domain knowledge.
- Concepts of retraining Embeddings based on, Intrinsic & Extrinsic evaluations
- Key takeaways – Building a better understanding of using word embeddings.

Part 3: Training & Designing Recurrent Neural Network | Multiclass Multilabel Text Classification (30min)

- How to finalise Neural Networks? RNN/CNN/Feed Forward.
- Not using Neural Networks as a silver bullet for the problem. Understanding the problem and re-designing the Pipeline Architecture.
    - Define Loss Functions & re-defining Evaluation Metrics.
    - An Ensemble approach to divide & Conquer the problem.
- How does one train – with code.
- Key Takeaway – Model Building & Solutioning for the problem dataset

Part 4: Hands on (60 min)

- Highlight main components as discussed in the Part 1,2,3
- Help attendees create their own custom word embeddings.

- Help attendees decide their model, hyper-parameters & fine-tuning.
- Experiment 1: Go through various embedding, single stacked. Compare results.
- Experiment 2: switch between single / ensemble NN. Compare results.

Learning Outcome

Key takeaways from this tutorial

- Understand what is Multi Class-Multi Label classification and how to solve it?
- Create & stack word embeddings.
- Approach to train a Neural Network that can solve MM Classification.

Target Audiences

Data Scientists, AI Enthusiasts, Managers, Engineers.