Weekly Report

WU Zihan

October 19, 2023

- Engaged in productive exchanges with Mehdi, sharing valuable data and code for our ongoing projects.
- Encountered challenges with the lack of 'ground truth', raising considerations for its impact on our research publication validity.
- I think to co-cluster the positional data along the time is a better idea.

Detailed Research in NLP

- **Theme classification** Studies [1] [2] demonstrate the efficacy of two-way clustering of documents and words, facilitating refined document and topic categorization. Notable datasets include "Yahoo_K5", "Classic3_3Odocs", and "Classic3_150docs".
- **Transfer Learning in NLP:** Research [3] highlights the role of co-clustering as a connecting bridge between labeled and unlabeled data, enabling effective topic classification beyond domain boundaries.
- **Information Dissemination:** Research [4] reveals the concept of overlapping co-clustering is leveraged for targeted information and news distribution, enhancing reach and relevance.

References: [1] Dhillon, 2001; [2] Dhillon et al., 2003; [3] Dai et al., 2007; [4] Sarkar & Srivastava, 2014.

Accelarating Multimodal Learning based on Co-clustering

- Recognized NLP's shift to multimodal models, merging diverse data types.
- Uncovered co-clustering's potential to enhance multimodal integration.
- Addressed training inefficiencies in contrastive learning.
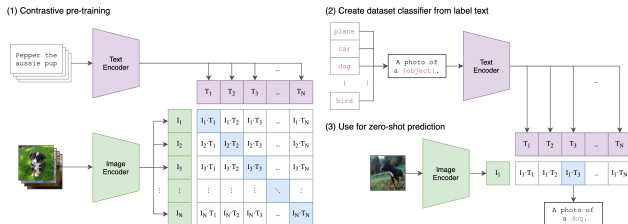- Suggested co-clustering to expedite training by reducing redundant computations.

Figure: CLIP [1]'s training process.

Reference: [1] Radford et al., 2021.