The document titled "Unsupervised multimodal learning for image-text relation classification in tweets" by Lin Sun et al. (2023) addresses the challenge of classifying the relationship between text and images in social media content, specifically tweets. The authors identify that manual labeling of image-text relations (ITR) is time-consuming and often leads to inconsistent labels due to disagreements among annotators. To overcome this, they propose an unsupervised learning method called ITR pseudo-labeling (ITRp), which uses clustering to generate pseudo-labels for training a classifier and finetuning encoders.

The ITRp method leverages pre-trained models (PTMs) for feature extraction and employs k-means clustering to create pseudo-labels that represent different types of ITR. These pseudo-labels are then used to train a binary classifier iteratively, improving the model's performance without the need for manual annotation. The authors evaluate their method on the ITR dataset and demonstrate that it achieves competitive results compared to supervised learning approaches, particularly when dealing with incorrect labels in the dataset.

The paper also discusses the importance of understanding ITR for improving website design, user experience, search engine ranking, and marketing strategies. It highlights the difficulty of obtaining reliable manual annotations for ITR, with inter-annotator agreement being low, especially for the text task. The authors' contribution includes identifying and correcting incorrectly labeled samples in the ITR dataset, which significantly impacted the performance of supervised learning models.

The ITRp method demonstrates the potential of unsupervised learning in addressing the challenges of multimodal data classification in tweets. The authors provide their code and data on a public GitHub repository, and their experimental results show that the ITRp method outperforms supervised models on a corrected test set, with improvements in F1 scores for both text and image tasks. The paper concludes by suggesting future work to explore the impact of noisy data on clustering performance and the development of larger, more diverse datasets for robust model evaluation.