

HyperLearn: A Distributed Approach for Representation Learning in Datasets with Many Modalities

Devanshu Arya, Stevan Rudinac & Marcel Worring



UNIVERSITEIT VAN AMSTERDAM

INTRODUCTION

Multimodal datasets contain an enormous amount of relational information which possesses a huge potential for learning in multimedia tasks.

Relations in datasets with many modalities can be efficiently represented using HyperLearn.



PROPOSED FRAMEWORK



EXPERIMENTAL SETUP

Evaluate performance and computational time of HyperLearn on real-world multimodal datasets.

GOAL: Learn representations for each entity within a modality from the intra and cross modal relations $\mathscr{R}()$.

| MATRIX COMPLETION O | N GRAPHS |
|-------------------------|----------|
| Dataset: MovieLens 100K | |
| Movies (M) | 1,682 |
| $ _{aara} ()$ | 042 |



CONTRIBUTIONS

A hypergraph-based framework for learning **complex higher-order** relationships in multimedia datasets

An unsupervised training approach capable of jointly modelling relations between the items of the **same modality**, as well as across different modalities.

Using HyperLearn, training can be parallelized easily across multiple GPUs, which makes it scalable to datasets with many modalities







In a Hypergraph, an edge is a connection between any number of vertices instead of pairwise connections

Mathematically, Hypergraph is represented by an adjacency Tensor instead of a matrix



Multi-Graph $\widetilde{A_{\theta}} = \sum_{jj'} \theta_{jj'} T_j(\mathbb{L}_{\theta}) A_{\theta}$ Convolution *j*,*j*′=0

 $\min_{A_1,..,A_K} \frac{1}{2} (tr \sum_{\theta=1}^{K} A_{\theta}^T \mathbb{L}_{\theta} A_{\theta}) + \lambda || \underline{\mathbf{X}} - (\mathbf{I} \times_1 A_1 \times_2 ... \times_K A_K) ||_F^2$ Cross-Modal Intra-Modal Loss Loss

Distributive Training Loss $Loss_{\theta} = \lambda ||X_{\theta} - A_{\theta}\Omega_{\theta}||_{F}^{2} + \frac{1}{2}tr(A_{\theta}^{T}\mathbb{L}_{\theta}A_{\theta})$

Tensor Unfolding $\Omega_{\theta} = A_1 \odot A_2 \odot .. \odot A_{\theta-1} \odot A_{\theta+1} .. \odot A_{\kappa}$

| 4,628,009 | 849,482 | 21,178 | 144 | 28,399 |
|-----------|---------|--------|-----|--------|
| | | | | |



(a) Vincent van Gogh – Olive Trees (b) Claude Monet – Marine View with a with Yellow Sky and Sun, 1889 Sunset, 1875

Van Gogh (a) and Monet (b) have many stylistic similarities, but their materialization is different. Capturing their similarities, differences and influences requires the ability to model higher-order relations.