Matrix Tri-Factorization with Manifold Regularizations for Zero-shot Learning
Xing Xu, Fumin Shen, Yang Yang, Dongxiang Zhang, Heng Tao Shen and Jingkuan Song
Center For Future Media & School of Computer Science and Engineering
University of Electronic Science and Technology of China, China

### Background
- **Zero-shot learning (ZSL)**
  - Two types of classes:
    - Seen: with labeled instances (for training only)
    - Unseen: without instances (for testing only)
  - Goal: Recognizing unseen object classes based on the knowledge learned from seen object classes during training.

### Proposed Method
- **Matrix Tri-Factorization with Manifold Regularization (MFMR)**
  - Learning the projection from visual features of seen classes
  - $\mathbf{X} = \mathbf{U} \times \mathbf{A} \times \mathbf{V}^T + \mathbf{R}(\mathbf{U}) + \mathbf{R}(\mathbf{V})$
    - $\mathbf{X} \in \mathbb{R}^{n \times m}$ is the projection, each $\mathbf{U}$ represents a visual feature cluster for each semantic embedding (e.g., attributes).
    - $\mathbf{A} \in \mathbb{R}^{m \times r}$ is the semantic embeddings (e.g., attributes).
    - $\mathbf{V} \in \mathbb{R}^{r \times s}$, each $\mathbf{V}$ represents an instance cluster for each seen class.
  - **Predicting the categories of unseen classes instances**
    - Simple prediction scheme (MFMR):
      $\mathbf{y} = \text{arg min}_i (\mathbf{U}^T \mathbf{x})$, $y_i \in \{1, \ldots, L\}$
    - Joint prediction scheme (MFMR-joint):Exploiting the manifold structure in unseen classes instances

### Experimental settings
- **Datasets**
  - AWA
  - CUB
  - SUN
  - CUB
  - SUN
  - SunHindi
  - Meta-objectives
  - BOW
  - HLP
  - INT

#### Results on "Conventional" Setting

<table>
<thead>
<tr>
<th>Method</th>
<th>AWA</th>
<th>CUB</th>
<th>SUN</th>
<th>SynC</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFMR</td>
<td>74.9</td>
<td>76.5</td>
<td>75.4</td>
<td>85.0</td>
<td>77.3</td>
</tr>
<tr>
<td>MFMR-joint</td>
<td>81.6</td>
<td>85.5</td>
<td>84.9</td>
<td>88.6</td>
<td>85.6</td>
</tr>
</tbody>
</table>

#### Results on "Generalized" Setting

<table>
<thead>
<tr>
<th>Method</th>
<th>AWA</th>
<th>CUB</th>
<th>SUN</th>
<th>SynC</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFMR</td>
<td>74.9</td>
<td>76.5</td>
<td>75.4</td>
<td>85.0</td>
<td>77.3</td>
</tr>
<tr>
<td>MFMR-joint</td>
<td>81.6</td>
<td>85.5</td>
<td>84.9</td>
<td>88.6</td>
<td>85.6</td>
</tr>
</tbody>
</table>

### Experiments
- **Evaluation tasks and metrics**
  - Conventional & Generalized ZSL Tasks: zero-shot classification & retrieval
  - Metrics: Mean Absolute Precision (MAP)
- **Compared methods**
  - DAP [CVPR'09], ALE [CVPR'13], ESZSL [ICML'15], TMV-HLP [TPAMI'15], SSE [ICCV'15], JSLE [CVPR'16], SynC [CVPR'16]

### Challenges in learning the projection
- The intrinsic manifold structure in the semantic embeddings of classes is not well explored.
- The projection shift problem exists due to the different distribution of seen and unseen classes.

### Knowledge transfer to unseen classes
- Learning a projection from the visual feature space to the semantic embedding space based on seen classes, and apply it to unseen classes.

### Details analysis
- The propose MFMR and MFMR-joint consistently perform the best on "generalized" setting.
- The propose MFMR and MFMR-joint predict more accurate results on "generalized" zero-shot classification task.