Relational Deep Learning: A Deep Latent Variable Model for Link Prediction

Hao Wang, Xingjian Shi, Dit-Yan Yeung
• Motivation
• Bayesian Deep Learning
• Relational Deep Learning
• Parameter Learning
• Experiments
• Conclusion
Motivation: Link Prediction

Social Network Analysis (e.g., prediction friendship in Facebook)
Motivation: Link Prediction

Document Networks (e.g., citation networks, co-author networks)
Motivation: Deep Latent Variable Models
Motivation: Deep Latent Variable Models

Stacked denoising autoencoders

Convolutional neural networks

Recurrent neural networks

Typically for i.i.d. data
• Motivation
• Bayesian Deep Learning
• Relational Deep Learning
• Parameter Learning
• Experiments
• Conclusion
Bayesian Deep Learning

Bayesian deep learning (BDL)
• Maximum a posteriori (MAP)
• Markov chain Monte Carlo (MCMC)
• Variational inference (VI)

Perception component
Content understanding
- Posts by users
- Text in articles

Task-Specific component
Target task
Link prediction

[ Wang et al. 2016 ]
# Bayesian Deep Learning

<table>
<thead>
<tr>
<th>Applications</th>
<th>Models</th>
<th>Hinge Variables</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recommender Systems</strong></td>
<td>CDL [Wang et al.]</td>
<td>{V}</td>
<td>MAP</td>
</tr>
<tr>
<td></td>
<td>Bayesian CDL [Wang et al.]</td>
<td>{V}</td>
<td>Gibbs Sampling</td>
</tr>
<tr>
<td></td>
<td>Marginalized CDL [Li et al.]</td>
<td>{V}</td>
<td>MAP</td>
</tr>
<tr>
<td></td>
<td>Symmetric CDL [Li et al.]</td>
<td>{V, U}</td>
<td>MAP</td>
</tr>
<tr>
<td></td>
<td>Collaborative Deep Ranking [Ying et al.]</td>
<td>{V}</td>
<td>MAP</td>
</tr>
<tr>
<td><strong>Topic Models</strong></td>
<td>Relational SDAE [Wang et al.]</td>
<td>{S}</td>
<td>MAP</td>
</tr>
<tr>
<td></td>
<td>DPFA-SBN [Gan et al.]</td>
<td>{X}</td>
<td>Hybrid MC</td>
</tr>
<tr>
<td></td>
<td>DPFA-RBM [Gan et al.]</td>
<td>{X}</td>
<td>Hybrid MC</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>Embed to Control [Watter et al.]</td>
<td>{z_t, z_{t+1}}</td>
<td>Variational Inference</td>
</tr>
</tbody>
</table>
A Principled Probabilistic Framework

Perception Component

Task-Specific Component

Perception Variables

Task Variables

Hinge Variables

[Wang et al. 2016]
• Motivation
• Bayesian Deep Learning
• **Relational Deep Learning**
• Parameter Learning
• Experiments
• Conclusion
Relational Deep Learning: Graphical Model

Perception component: relational and deep representation learning

Task-specific component: link prediction
Stacked Denoising Autoencoders (SDAE)

SDAE solves the following optimization problem:

\[
\min_{\{W_l\},\{b_l\}} \|X_c - X_L\|_F^2 + \lambda \sum_l \|W_l\|_F^2,
\]

where \(\lambda\) is a regularization parameter and \(\|\cdot\|_F\) denotes the Frobenius norm.

[Vincent et al. 2010]
Probabilistic SDAE

Graphical model:

Generative process:

\[
W_{l,*n} \sim \mathcal{N}(0, \lambda_w^{-1}I_{K_l})
\]

\[
b_l \sim \mathcal{N}(0, \lambda_w^{-1}I_{K_l})
\]

\[
X_{l,j*} \sim \mathcal{N}(\sigma(X_{l-1,j*}W_l + b_l), \lambda_s^{-1}I_{K_l})
\]

\[
X_{c,j*} \sim \mathcal{N}(X_{L,j*}, \lambda_n^{-1}I_B)
\]

Generalized SDAE

Notation:

- \(X_0\): corrupted input
- \(X_c\): clean input
- \(W^+\): weights and biases

[Wang et al. 2015]
Relational Deep Learning

Probabilistic SDAE

Modeling relation among nodes
Network of Probabilistic SDAE

Many interconnected probabilistic SDAEs with shared weights
• Motivation
• Bayesian Deep Learning
• Relational Deep Learning
• Parameter Learning
• Experiments
• Conclusion
MAP Inference

maximizing the posterior probability is equivalent to maximizing the joint log-likelihood

$$\mathcal{L} = -\frac{\lambda_w}{2} \sum_l (\|W_l\|_F^2 + \|b_l\|_2^2)$$

$$-\frac{\lambda_p}{2} \sum_i \|\phi_i - X_{L_{i*}}^T \|_2^2 - \frac{\lambda_n}{2} \sum_i \|X_{L,i*} - X_{c,i*}\|_2^2$$

$$-\frac{\lambda_s}{2} \sum_l \sum_i \|\sigma(X_{l-1,i*} W_l + b_l) - X_{l,i*}\|_2^2$$

$$-\frac{\lambda_e}{2} \|\eta\|_2^2 + \sum_{l_i,i' = 1} \log \sigma(\eta^T (\phi_i \circ \phi_{i'})).$$
MAP Inference

Prior (regularization) for link prediction parameters, weights, and biases

\[ \mathcal{L} = -\frac{\lambda_w}{2} \sum_l (\|W_l\|_F^2 + \|b_l\|_2^2) \]

\[ -\frac{\lambda_p}{2} \sum_i \|\phi_i - X_{L,i*}^T\|_2^2 - \frac{\lambda_n}{2} \sum_i \|X_{L,i*} - X_{C,i*}\|_2^2 \]

\[ -\frac{\lambda_s}{2} \sum_l \sum_i \|\sigma(X_{l-1,i*} W_l + b_l) - X_{l,i*}\|_2^2 \]

\[ -\frac{\lambda_e}{2} \|\eta\|_2^2 + \sum_{l_{i,i'=1}} \log \sigma(\eta^T (\phi_i \circ \phi_{i'})) \]
MAP Inference

Generating node features from content representation with Gaussian offset

\[
\mathcal{L} = - \frac{\lambda_w}{2} \sum_l (\|W_l\|_F^2 + \|b_l\|_2^2) \\
- \frac{\lambda_p}{2} \sum_i \|\phi_i - X_{L/2,i*}^T\|_2^2 \\
- \frac{\lambda_n}{2} \sum_i \|X_{L,i*} - X_{c,i*}\|_2^2 \\
- \frac{\lambda_s}{2} \sum_l \sum_i \|\sigma(X_{l-1,i*}W_l + b_l) - X_{l,i*}\|_2^2 \\
- \frac{\lambda_e}{2} \|\eta\|_2^2 + \sum_{l_i,i'=1} \log \sigma(\eta^T(\phi_i \circ \phi_{i'})).
\]
MAP Inference

‘Generating’ clean input from the output of probabilistic SDAE with Gaussian offset

\[
\mathcal{L} = -\frac{\lambda_w}{2} \sum_l (\|W_l\|_F^2 + \|b_l\|_2^2) \\
- \frac{\lambda_p}{2} \sum_i \| \phi_i - X_{L, i^*, i}^T \|^2_2 \\
- \frac{\lambda_n}{2} \sum_i \| X_{L, i^*} - X_{c, i^*} \|^2_2 \\
- \frac{\lambda_s}{2} \sum_l \sum_i \| \sigma(X_{l-1, i^*} W_l + b_l) - X_{l, i^*} \|^2_2 \\
- \frac{\lambda_e}{2} \| \eta \|^2_2 + \sum_{l_i, i'_1=1} \log \sigma(\eta^T (\phi_i \circ \phi_{i'}).) 
\]
Generating the input of Layer $l$ from the output of Layer $l-1$ with Gaussian offset

$$\mathcal{L} = -\frac{\lambda_w}{2} \sum_l (\|W_l\|_F^2 + \|b_l\|_2^2)$$

$$-\frac{\lambda_p}{2} \sum_i \|\phi_i - X_{L/2,i,*}^T\|_2^2 - \frac{\lambda_n}{2} \sum_i \|X_{L,i,*} - X_{c,i,*}\|_2^2$$

$$-\frac{\lambda_s}{2} \sum_l \sum_i \|\sigma(X_{l-1,i,*}W_l + b_l) - X_{l,i,*}\|_2^2$$

$$-\frac{\lambda_e}{2} \|\eta\|_2^2 + \sum_{l_i,i' = 1} \log \sigma(\eta^T(\phi_i \circ \phi_{i'})).$$
MAP Inference

Generating links from Bernoulli distributions parameterized by $\eta$ and $\phi$

$$
\mathcal{L} = -\frac{\lambda_w}{2} \sum_l (\| W_l \|_F^2 + \| b_l \|_2^2) \\
- \frac{\lambda_p}{2} \sum_i \| \phi_i - X_{l,i}^{T,L,i*} \|_2^2 - \frac{\lambda_n}{2} \sum_i \| X_{L,i*} - X_{c,i*} \|_2^2 \\
- \frac{\lambda_s}{2} \sum_l \sum_i \| \sigma(X_{l-1,i*} W_l + b_l) - X_{l,i*} \|_2^2 \\
- \frac{\lambda_e}{2} \| \eta \|_2^2 + \sum_{l_i,i' = 1} \log \sigma(\eta^T (\phi_i \circ \phi_{i'}))$$
Bayesian Treatment:
Generalized Variational Inference

\[ \log q_j^*(\mathbf{Z}_j) = E_{i \neq j} [\log p(\mathbf{X}_0, \mathbf{X}_c, \mathbf{Z})] + const \]

- \( q_1(\mathbf{Z}_1) = q(\mathbf{W}^+) \): Variational distributions for weights/biases.
- \( q_2(\mathbf{Z}_2) = q(\phi_i) \): Variational distributions for generated node features.
- \( q_3(\mathbf{Z}_3) = q(\eta) \): Variational distributions for parameters of the link prediction model.
- \( q_4(\mathbf{Z}_4) = q(\zeta) \): Variational parameters to approximate the sigmoid function.

Use Laplace approximation rather than variational inference for weights/biases.
Example: Updating $\phi$ as a Product of Gaussians

Update $\phi$ for node $i$ as a product of two Gaussians

$$q(\phi_i | X_{0,i*}) \approx \mathcal{N}(\phi_i | \mu_i, \Sigma_i)$$

$$\mu_i = \Sigma_i (S_i^{-1} m_i + S_i'^{-1} m_i')$$

$$\Sigma_i^{-1} = S_i^{-1} + S_i'^{-1}.$$
• Motivation
• Bayesian Deep Learning
• Relational Deep Learning
• Parameter Learning
• Experiments
• Conclusion
Experiments: Settings

Document Networks (e.g., citation networks)

<table>
<thead>
<tr>
<th>datasets</th>
<th># nodes</th>
<th># links</th>
</tr>
</thead>
<tbody>
<tr>
<td>citeulike-a</td>
<td>16,980</td>
<td>44,709</td>
</tr>
<tr>
<td>citeulike-t</td>
<td>25,975</td>
<td>32,565</td>
</tr>
<tr>
<td>arXiv</td>
<td>27,770</td>
<td>352,807</td>
</tr>
</tbody>
</table>
Experiments: Link Rank and AUC

Figure 2: Link rank and AUC of compared models for citeulike-a. A 2-layer RDL is used.

Link rank: how high our predicted links rank in the ground truth
AUC: area under curve
Experiments: Link Rank and AUC

Figure 3: Link rank and AUC of compared models for citeulike-t. A 2-layer RDL is used.

Link rank: how high our predicted links rank in the ground truth
AUC: area under curve
Experiments: Link Rank and AUC

Figure 4: Link rank and AUC of compared models for arXiv. A 2-layer RDL is used.

Link rank: how high our predicted links rank in the ground truth
AUC: area under curve
Experiments: RDL Variants

Link rank of baselines (the first 3 columns) and RDL variants (the last 4 columns) on three datasets (L = 4)

<table>
<thead>
<tr>
<th>Method</th>
<th>VAE+BLR</th>
<th>VFAE+BLR</th>
<th>SDAE+BLR</th>
<th>MAPRDL</th>
<th>BSDAE1+BLR</th>
<th>BSDAE2+BLR</th>
<th>BayesRDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>citeulike-a</td>
<td>980.81</td>
<td>960.15</td>
<td>992.48</td>
<td>495.97</td>
<td>849.02</td>
<td>761.57</td>
<td>473.59</td>
</tr>
<tr>
<td>citeulike-t</td>
<td>1599.62</td>
<td>1531.16</td>
<td>1356.85</td>
<td>951.31</td>
<td>1341.15</td>
<td>1310.12</td>
<td>911.31</td>
</tr>
<tr>
<td>arXiv</td>
<td>3367.25</td>
<td>3316.29</td>
<td>2916.18</td>
<td>2028.72</td>
<td>2947.79</td>
<td>2708.17</td>
<td>1982.84</td>
</tr>
</tbody>
</table>

VAE: Variational Autoencoder
VRAE: Variational Fair Autoencoder
BLR: Bayesian Logistic Regression
BSDAE1: Bayesian treatment of probabilistic SDAE (mean only)
BSDAE2: Bayesian treatment of probabilistic SDAE (mean and variance)
MAPRDL: RDL with MAP inference
BayesRDL: RDL with full Bayesian treatment
Experiments: Depth

Performance of RDL with different number of layers (MAP)

<table>
<thead>
<tr>
<th></th>
<th>Link Rank</th>
<th></th>
<th>AUC</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RDL-1</td>
<td>RDL-2</td>
<td>RDL-3</td>
<td>RDL-1</td>
<td>RDL-2</td>
<td>RDL-3</td>
</tr>
<tr>
<td>citeulike-a</td>
<td>825.74</td>
<td>495.97</td>
<td>488.41</td>
<td>0.939</td>
<td>0.964</td>
<td>0.963</td>
</tr>
<tr>
<td>citeulike-t</td>
<td>2060.17</td>
<td>951.31</td>
<td>912.43</td>
<td>0.894</td>
<td>0.954</td>
<td>0.955</td>
</tr>
<tr>
<td>arXiv</td>
<td>5241.97</td>
<td><strong>2080.72</strong></td>
<td>2730.08</td>
<td>0.755</td>
<td><strong>0.905</strong></td>
<td>0.855</td>
</tr>
</tbody>
</table>

Performance of RDL with different number of layers (Bayesian treatment)

<table>
<thead>
<tr>
<th></th>
<th>Link Rank</th>
<th></th>
<th>AUC</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RDL-1</td>
<td>RDL-2</td>
<td>RDL-3</td>
<td>RDL-1</td>
<td>RDL-2</td>
<td>RDL-3</td>
</tr>
<tr>
<td>citeulike-a</td>
<td>789.85</td>
<td>473.59</td>
<td><strong>471.47</strong></td>
<td>0.946</td>
<td><strong>0.971</strong></td>
<td>0.970</td>
</tr>
<tr>
<td>citeulike-t</td>
<td>1904.83</td>
<td>911.31</td>
<td><strong>867.78</strong></td>
<td>0.906</td>
<td>0.956</td>
<td><strong>0.960</strong></td>
</tr>
<tr>
<td>arXiv</td>
<td>4965.01</td>
<td><strong>1982.84</strong></td>
<td>2612.12</td>
<td>0.801</td>
<td><strong>0.914</strong></td>
<td>0.866</td>
</tr>
</tbody>
</table>
Case Study: RDL and RTM

t-SNE visualization of latent factors learned by RDL (left) and RTM (right).

Target article:
From DNA sequence to transcriptional behaviour: a quantitative approach

🌟 (red): articles with links to the target article
〇 (blue): articles without links to the target article
Case Study: RDL

Articles written in **German**, which are rare in the datasets

Some **bestselling books**
The 4-Hour Work Week
Mary Bell’s Complete Dehydrator Cookbook

**t-SNE visualization of latent factors learned by RDL.**

**Target article:**
From DNA sequence to transcriptional behaviour: a quantitative approach
Case Study: RDL and gRTM

Top 10 link predictions made by gRTM and RDL for two articles from citeulike-a

<table>
<thead>
<tr>
<th>Query: Object class recognition by unsupervised scale-invariant learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layered depth images</td>
</tr>
<tr>
<td>Using spin images for efficient object recognition in cluttered 3D scenes</td>
</tr>
<tr>
<td>Snakes: active contour models</td>
</tr>
<tr>
<td>Visual learning and recognition of 3-D objects from appearance</td>
</tr>
<tr>
<td>Contextual priming for object detection</td>
</tr>
<tr>
<td>Visual categorization with bags of keypoints</td>
</tr>
<tr>
<td>Non-parametric model for background subtraction</td>
</tr>
<tr>
<td>Alignment by maximization of mutual information</td>
</tr>
<tr>
<td><strong>Rapid object detection using a boosted cascade of simple features</strong></td>
</tr>
<tr>
<td><strong>W4: real-time surveillance of people and their activities</strong></td>
</tr>
<tr>
<td>Distinctive image features from scale-invariant keypoints</td>
</tr>
<tr>
<td>visual learning and recognition of 3-D objects from appearance</td>
</tr>
<tr>
<td><strong>Object recognition with features inspired by visual cortex</strong></td>
</tr>
<tr>
<td>Unsupervised learning of models for recognition</td>
</tr>
<tr>
<td><strong>Robust object recognition with cortex-like mechanisms</strong></td>
</tr>
<tr>
<td>Generative versus discriminative methods for object recognition</td>
</tr>
<tr>
<td>Using spin images for efficient object recognition in cluttered 3D scenes</td>
</tr>
<tr>
<td><strong>Learning generative visual models from few training examples</strong></td>
</tr>
<tr>
<td>3D object modeling and recognition using affine-invariant patches</td>
</tr>
<tr>
<td>A Bayesian approach to unsupervised one-shot learning of object categories</td>
</tr>
</tbody>
</table>

**Key Concepts**
- Object recognition
- Unsupervised learning
- Scale-invariant learning
Case Study: RDL and gRTM

Top 10 link predictions made by gRTM and RDL for two articles from citeulike-a

<table>
<thead>
<tr>
<th>gRTM</th>
<th>Query: SCOP database in 2004: refinements integrate structure and sequence family data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pfam: multiple sequence alignments and HMM-profiles of protein domains</td>
</tr>
<tr>
<td></td>
<td><strong>Structure, function and evolution of multidomain proteins</strong></td>
</tr>
<tr>
<td></td>
<td>Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB</td>
</tr>
<tr>
<td></td>
<td>Nature of the protein universe</td>
</tr>
<tr>
<td></td>
<td>The CATH domain structure database and related resources</td>
</tr>
<tr>
<td></td>
<td>Phylogenetic classification of short environmental DNA fragments</td>
</tr>
<tr>
<td></td>
<td>The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes</td>
</tr>
<tr>
<td></td>
<td>LGA: a method for finding 3D similarities in protein structures</td>
</tr>
<tr>
<td></td>
<td>Amino acid substitution matrices from protein blocks</td>
</tr>
<tr>
<td></td>
<td>Multiple protein sequence alignment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RDL</th>
<th>The universal protein resource (UniProt)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>E-MSD: an integrated data resource for bioinformatics</strong></td>
</tr>
<tr>
<td></td>
<td>Gene3D: comprehensive structural and functional annotation of genomes</td>
</tr>
<tr>
<td></td>
<td>The universal protein resource (UniProt) in 2010</td>
</tr>
<tr>
<td></td>
<td><strong>Gene3D: modelling protein structure, function and evolution</strong></td>
</tr>
<tr>
<td></td>
<td>The universal protein resource (UniProt): an expanding universe of protein information</td>
</tr>
<tr>
<td></td>
<td>Pfam: clans, web tools and services</td>
</tr>
<tr>
<td></td>
<td>The Pfam protein families database</td>
</tr>
<tr>
<td></td>
<td>The protein data bank</td>
</tr>
<tr>
<td></td>
<td>SCOP: a structural classification of proteins database</td>
</tr>
</tbody>
</table>

Key Concepts
- Protein structures
- Protein databases
• Motivation
• Bayesian Deep Learning
• Relational Deep Learning
• Parameter Learning
• Experiments
• Conclusion
Conclusion

- First Bayesian DL model for link prediction
- Joint Bayesian DL is beneficial
- Significant improvement on the state of the art
- RDL as representation learning
Future Work

• Multi-relational data (co-author & citation networks)
  • Boost predictive performance
  • Discover relationship between different networks

• GVI for other neural nets (CNN/RNN) and BayesNets
  • pSDAE + link prediction
  • pCNN + recommendation
  • pRNN + community detection

• Replace probabilistic SDAE with other Bayesian neural nets
  • Variational autoencoders
  • Natural-parameter networks
THANK YOU