Intro to representational learning

MLRG
February 2019
Data $\rightarrow$ vector

- Handwritten scans
- Seismic scans
- Cell imaging
- Raw images
- Astronomy
- Stock market

Diagram:
- Paris
- SeaWorld
- porpoise
- dolphin

(Quora)
Why representational learning

It looks pretty
Supervised learning
Unsupervised learning
Data mining
It looks pretty

Mapping Word Embeddings with Word2vec: Enhancing Natural Language Processing with Semantic and Syntactic Relationships between Word Vectors

Sam Liebman (2018)
It looks pretty

Pennington, Manning, Socher, 2014
CNN layers

Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps
Simonyan, Vedaldi, Zisserman 2014
Supervised learning

“The cat and the hat”

<table>
<thead>
<tr>
<th>Input</th>
<th>Hidden Layers</th>
<th>Part-of-Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>bat</td>
<td></td>
<td>1 Noun</td>
</tr>
<tr>
<td>cat</td>
<td></td>
<td>0 Verb</td>
</tr>
<tr>
<td>fat</td>
<td></td>
<td>0 Adj.</td>
</tr>
<tr>
<td>hat</td>
<td></td>
<td>0 Adv.</td>
</tr>
<tr>
<td>mat</td>
<td></td>
<td>0 Prep.</td>
</tr>
<tr>
<td>sat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>that</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supervised learning

"The cat and the hat"

\( X_{\text{cat}} \)

Input

Hidden layers

Part-of-speech

1 Noun
0 Verb
0 Adj.
0 Adv.
0 Prep.

fat

bat_c\text{at}

sat

that

hat_mat
NLP named entity recognition

<table>
<thead>
<tr>
<th>System</th>
<th>Dev</th>
<th>Test</th>
<th>MUC7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>90.03</td>
<td>84.39</td>
<td>67.48</td>
</tr>
<tr>
<td>Baseline+Nonlocal</td>
<td>91.91</td>
<td>86.52</td>
<td>71.80</td>
</tr>
<tr>
<td>HLBL 100-dim</td>
<td>92.00</td>
<td>88.13</td>
<td>75.25</td>
</tr>
<tr>
<td>Gazetteers</td>
<td>92.09</td>
<td>87.36</td>
<td>77.76</td>
</tr>
<tr>
<td>C&amp;W 50-dim</td>
<td>92.27</td>
<td>87.93</td>
<td>75.74</td>
</tr>
<tr>
<td>Brown, 1000 clusters</td>
<td>92.32</td>
<td>88.52</td>
<td>78.84</td>
</tr>
<tr>
<td>C&amp;W 200-dim</td>
<td>92.46</td>
<td>87.96</td>
<td>75.51</td>
</tr>
<tr>
<td>C&amp;W+HLBL</td>
<td>92.52</td>
<td>88.56</td>
<td>78.64</td>
</tr>
<tr>
<td>Brown+HLBL</td>
<td>92.56</td>
<td>88.93</td>
<td>77.85</td>
</tr>
<tr>
<td>Brown+C&amp;W</td>
<td>92.79</td>
<td>89.31</td>
<td>80.13</td>
</tr>
<tr>
<td>HLBL+Gaz</td>
<td>92.91</td>
<td>89.35</td>
<td>79.29</td>
</tr>
<tr>
<td>C&amp;W+Gaz</td>
<td>92.98</td>
<td>88.88</td>
<td>81.44</td>
</tr>
<tr>
<td>Brown+Gaz</td>
<td>93.25</td>
<td>89.41</td>
<td>82.71</td>
</tr>
<tr>
<td>Lin and Wu (2009), 3.4B</td>
<td>-</td>
<td>88.44</td>
<td>-</td>
</tr>
<tr>
<td>Ando and Zhang (2005), 27M</td>
<td>93.15</td>
<td>89.31</td>
<td>-</td>
</tr>
<tr>
<td>Suzuki and Isozaki (2008), 37M</td>
<td>93.66</td>
<td>89.36</td>
<td>-</td>
</tr>
<tr>
<td>Suzuki and Isozaki (2008), 1B</td>
<td>94.48</td>
<td>89.92</td>
<td>-</td>
</tr>
<tr>
<td>All (Brown+C&amp;W+HLBL+Gaz), 37M</td>
<td>93.17</td>
<td>90.04</td>
<td>82.50</td>
</tr>
<tr>
<td>All+Nonlocal, 37M</td>
<td>93.95</td>
<td>90.36</td>
<td>84.15</td>
</tr>
<tr>
<td>Lin and Wu (2009), 700B</td>
<td>-</td>
<td>90.90</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3: Final NER F1 results, showing the cumulative effect of adding word representations, non-local features, and gazetteers to the baseline. To speed up training, in combined experiments (C&W plus another word representation), we used the 50-dimensional C&W embeddings, not the 200-dimensional ones. In the last section, we show how many unlabeled words were used.

Word representations: A simple and general method for semi-supervised learning
Turian, Ratinov, Bengio, 2010
Clustering

Figure 2: Comparison of adversarial and variational autoencoder on MNIST. The hidden code $z$ of the hold-out images for an adversarial autoencoder fit to (a) a 2-D Gaussian and (b) a mixture of 10 2-D Gaussians. Each color represents the associated label. Same for variational autoencoder with (c) a 2-D gaussian and (d) a mixture of 10 2-D Gaussians. (e) Images generated by uniformly sampling the Gaussian percentiles along each hidden code dimension $z$ in the 2-D Gaussian adversarial autoencoder.
Disambiguation and style transfer

A Neural Algorithm of Artistic Style
Gatys, Ecker, and Bethge, 2015

Artistic Style Transfer
Doukkali, 2018
Past methods

Spectral methods
SIFT / HOG
An SDP approach
Also called manifold learning

Also called manifold learning

Scale Invariant Feature Transform (SIFT)

Distinctive Image Features from Scale-Invariant Keypoints
Lowe, 2004
Histogram of oriented gradients

Method of and apparatus for pattern recognition
McConnell 1986

Histograms of Oriented Gradients for Human Detection
Dalal and Triggs 2005

Video Processing Algorithms for Detection of Pedestrians
Piniarski, Pawlowski, Dabrowski, 2015
Maximum variance unfolding (SDP approach)

An Introduction to Nonlinear Dimensionality Reduction by Maximum Variance Unfolding
Weinberger and Saul 2006
Neural network methods

t-SNE
Word2vec
Transfer learning
Autoencoders
t-Distributed Stochastic Neighbor Embedding (t-SNE)

Visualizing High-Dimensional Data Using t-SNE
van der Maaten and Hinton 2008
Word embeddings

• Low-rank SVD ...
  • ... on document-word counts (LSI, Deerwester et al 1990)
  • ... on cooccurrence matrix (HAL, Lund / Burgess 1996)
  • ... on PPMI matrix (Bullinaria / Levy 2007)

• Shallow neural network
  • Bengio 2003, Collobert/Weston 2008, Mikolov et al 2013 (w2v)

• Least squares
  • GloVe (Pennington et al 2014)

• ... many unlisted ...
Similar technique on other data

Harer et al 2018
Transfer Learning for Natural Language Processing
Radhakrishnan, 2018

Transfer learning

DeepImageFeaturizer

Classifier

Chihuahua

GIANT PANDA 0.9
RED PANDA 0.05
RACCOON 0.01
...
Autoencoders

Understanding Autoencoders -- Unsupervised Learning Technique
Mishra, 2018

Recommendation systems with deep autoencoders
Kashikar
<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 4</td>
<td>Intro</td>
<td>Yifan</td>
</tr>
<tr>
<td>Feb 11</td>
<td>Artistic style transfer</td>
<td>amir</td>
</tr>
<tr>
<td>Feb 18 (Holiday)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb 25</td>
<td>GANS</td>
<td>aaron</td>
</tr>
<tr>
<td>Mar 4</td>
<td>autoencoders</td>
<td>vaden</td>
</tr>
<tr>
<td>Mar 11</td>
<td>Convolutional graph embeddings</td>
<td>cathy</td>
</tr>
<tr>
<td>Mar 18</td>
<td>Manifold learning</td>
<td>wilder</td>
</tr>
<tr>
<td>Mar 25</td>
<td>Graph and point cloud embeddings</td>
<td>marjan</td>
</tr>
<tr>
<td>Apr 1</td>
<td>Metric learning</td>
<td>mehrdad</td>
</tr>
<tr>
<td>Apr 8</td>
<td>disentanglement</td>
<td>Michael p</td>
</tr>
<tr>
<td>Apr 15 (AISTATS)</td>
<td>Dictionary learning</td>
<td>yihan</td>
</tr>
</tbody>
</table>

**Topic Options:**
- Sparse coding / dictionary learning
- Disentanglement
- Convolutional graph embeddings
- Style transfer
- T-SNE
- Variational autoencoders
- GANs
- Other types of embeddings (graphs, point clouds, video...)
- Metric learning (learning similarities)
- <other>