IIT – Madras: AIDB Lab

Presented by:

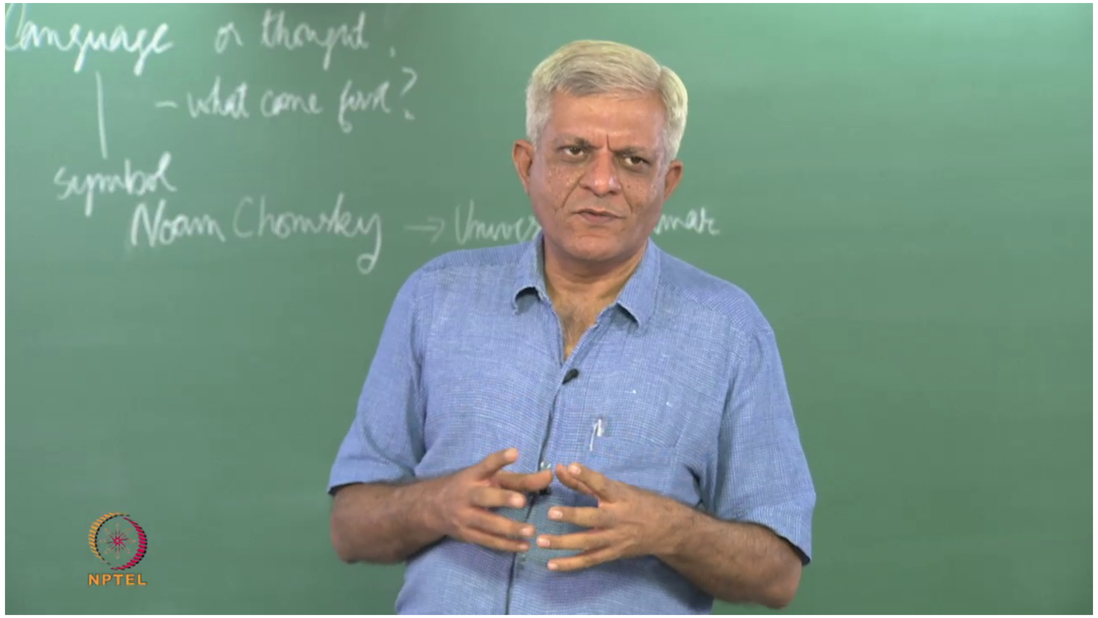
Pravar Mahajan & Sankeerth V S

Artificial Intelligence & Database Lab

Faculty	Research Area
Deepak Khemani	Artificial intelligence, Knowledge-based systems, Case-based reasoning, Knowledge representation, Planning, Logic, Natural language processing
NS Narayanswamy	Analysis of algorithms, Complexity theory, Artificial Intelligence
Sayan Raju	Graph Indexing, Graph Mining, Trajectory Analytics, Bioinformatics.
P Sreenivasa Kumar	Database systems, Semi-structured data and XML, Data mining, Graph algorithms, Parallel computing
Sutanu Chakraborti	Natural Language Processing,Information retrieval, Memory-based reasoning, Machine learning
B. Ravindran	Reinforcement learning, Computational Neuroscience, Data and text mining,Robotics,Social Network Analysis

Deepak Khemani

- **Ph.D.** in Computer Science, IIT Bombay, 1989
- **Research Focus:** Memory Based Reasoning, Knowledge Representation and Reasoning, Planning and Constraint Satisfaction, Qualitative Reasoning, Natural Language Generation.
- **Teaching:** Artificial Intelligence (NPTEL), Planning and Constraint Satisfaction, Memory Based Reasoning in AI, and Knowledge Representation and



Balaraman Ravindran

- **Ph.D.** in Computer Science, University of Massachusetts
- **Research Focus:** Reinforcement learning, Computational Neuroscience, Data and text mining,Robotics,Social Network Analysis
- **Teaching:** Topics in Reinforcement Learning(NPTEL), Introduction to Robotics, Natural Language Processing

Correlational Neural Networks (CorrNet)

- Motivation: Recent interest in learning common representation for multiple views of the data.

  - Reconstruction of Missing view: Consider a Movie Clip(audio, video, subtitles). Learned common representations can be used to train a model to reconstruct the subtitles even when only audio and video are available.

  - Transfer learning: Profanity detector trained on movie subtitles needs to detect profanities in a movie clip for which only video is available. If common representation is available for different views, then such detectors or classifiers can be trained by computing this common representation from the relevant view.

Correlational Neural Networks (CorrNet)

- Matching corresponding items across views: Items from one view (names written using the script of one language) need to be matched to their corresponding items from another view (names written using the script of another language).

- Improving single-view performance by using data from other views: Say, we are interested in learning word representations for a language. If we have access to translations of these words in another language, then these translations can provide some context for disambiguation, which can lead to learning better word representations.

The Model..

- Given a data $Z = \{\mathbf{z}_i\}_{i=1}^N$ with two views X and Y.
- Each data point $\mathbf{z}_i$ can represe $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i)$ $\mathbf{x}_i \in \mathbb{R}^{d_1}$ wh $\mathbf{y}_i \in \mathbb{R}^{d_2}$
  and .
- We are interested in learning two functions h_X and h_Y, such that $h_X(\mathbf{x}_i) \in \mathbb{R}^k$ and $h_Y(\mathbf{y}_i) \in \mathbb{R}^k$ $\mathbf{x}_i$ and $\mathbf{y}_i$ in $(\mathbb{R}^k)$

1. $h_X(\mathbf{x}_i)$ and $h_Y(\mathbf{y}_i)$ should be highly correlated.
2. It should be possible to reconstruct $\mathbf{y}_i$ from $\mathbf{x}_i$ (through $h_X(\mathbf{x}_i)$) and vice versa.

- Given $\mathbf{z} = (\mathbf{x}, \mathbf{y})$, we compute the encoded expression as

$$h(\mathbf{z}) = f(\mathbf{W}\mathbf{x} + \mathbf{V}\mathbf{y} + \mathbf{b})$$

where $\mathbf{W}$ is a $k \times d_1$ projection matrix, $\mathbf{V}$ is a $k \times d_2$ projection matrix, and $\mathbf{b}$ is a $k \times 1$ bias vector.

The Model..

- The Output layer tries to reconstruct **z** from the hidden

l $\mathbf{z}' = g([\mathbf{W}'h(\mathbf{z}), \mathbf{V}'h(\mathbf{z})] + \mathbf{b}')$

where $\mathbf{W}'$ is a $d_1 \times k$ reconstruction matrix, $\mathbf{V}'$ is a $d_2 \times k$ reconstruction matrix, and $\mathbf{b}'$ is a $(d_1 + d_2) \times 1$ output bias vector.

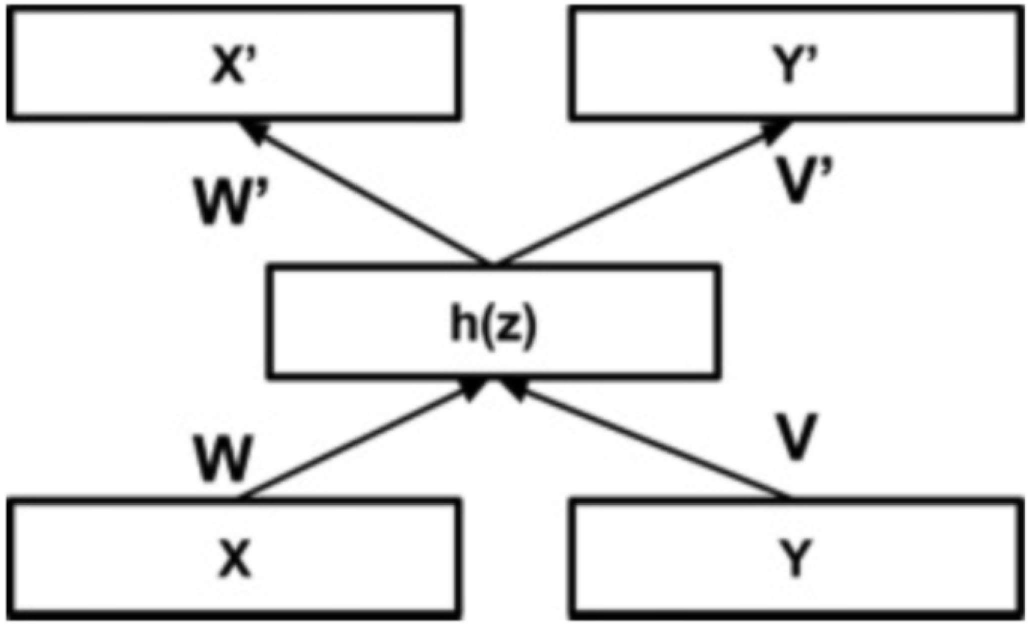
- function has to be

$$\mathcal{J}_{\mathcal{Z}}(\theta) = \sum_{i=1}^{N} (L(\mathbf{z}_i, g(h(\mathbf{z}_i))) + L(\mathbf{z}_i, g(h(\mathbf{x}_i))) + L(\mathbf{z}_i, g(h(\mathbf{y}_i)))) - \lambda \operatorname{corr}(h(X), h(Y))$$

w $$\operatorname{corr}(h(X), h(Y)) = \frac{\sum_{i=1}^{N} (h(\mathbf{x}_i) - \overline{h(X)})(h(\mathbf{y}_i) - \overline{h(Y)})}{\sqrt{\sum_{i=1}^{N} (h(\mathbf{x}_i) - \overline{h(X)})^2 \sum_{i=1}^{N} (h(\mathbf{y}_i) - \overline{h(Y)})^2}}$$

The Model..

- The architecture of the model can be viewed as,



Bridge CorrNets

- Motivation: Consider the problem of captioning images.
- Publicly available datasets contain images and their corresponding English notations.
- Hard to find similar datasets for languages such as Russian, Dutch, Hindi, Urdu, etc.
- Parallel dataset available only between English (pivotal language) and non-pivotal languages.
- How do we train our model to solve this problem in languages other than English?

Bridge CorrNets

- Learn aligned representations across multiple views using a pivotal view.
- Given only the parallel data between each of the n-views and the pivotal view.
- Downstream applications:
  - Transfer learning between languages L1, L2, L3 .. Ln using a pivotal language L
  - Cross modal access between images and a language L1 using pivotal language L

The Model..

- Based on the extension of the CorrNet model
- Let the views be denote $V_1, V_2, ..., V_M$ $d_1, d_2, ..., d_M$ et be their dimensions.
- Training dat $\mathcal{Z} = \{z^i\}_{i=1}^N$ with each $z^i = (v_j^i, v_M^i)$
- For each view, the encoding $h_{V_j}(v_j) = f(W_j v_j + b)$
- $W_j \in \mathbb{R}^{k \times d_j}$ is the encoder m V_j, $b \in \mathbb{R}^k$ is the common bias
- H n of the concatenated training instance: $h_Z(z) = f(W_j v_j + W_M v_M + b)$

The Model..

- Decoder corresponding to each $g_{V_j}(h) = p(W_j'h + c_j)$
- To train, the following objective function is minimized:

$$\mathcal{J}_{\mathcal{Z}}(\theta) = \sum_{i=1}^{N} L(z^i, g(h(z^i))) + \sum_{i=1}^{N} L(z^i, g(h(v^i_{l(i)}))) + \sum_{i=1}^{N} L(z^i, g(h(v^i_{M}))) - \lambda \operatorname{corr}(h(V_{l(i)}), h(V_M))$$

Corr-Net	Bridge Corr-Net
Sarath Chandar *University of Montreal*	**Janarthanan Rajendran** *IIT Madras,India*
Mitesh M. Khapra *IBM Research India*	**Mithesh M Khapra** *IBM Research India*
Hugo Larochelle *University of Sherbrooke*	**Sarath Chandar** *University of Montreal*
Balaraman Ravindran *Indian Institute of Technology, Madras*	**Balaraman Ravindran** *Indian Institute of Technology, Madras*

- The Corr-Net paper was selected as **one of the best 3 papers** of last year internal to **IBM Research India**.
- The Bridge Corr-Net paper was mentioned in an article on **top 25 AI breakthroughs** of 2015 by the **Future of Life Institute**.