deep learning and applications in non-cognitive domains

part 3

truyen tran
deakin university
truyen.tran@deakin.edu.au
prada-research.net/~truyen

AusDM'16, Canberra, Dec 7th 2016

http://www.tocci.com/wp-content/uploads/2015/06/150623_logotrends_feature.jpg

PART III: ADVANCED TOPICS

Unsupervised learning

Complex domain structures: Relations (explicit & implicit), graphs & tensors

Memory, attention

How to position ourselves

UNSUPERVISED LEARNING

WHY NEURAL UNSUPERVISED LEARNING?

Motivation: Humans mainly learn by exploring without clear instructions and labelling

Representational richness:

- FFN are functional approximator
- RNN are program approximator, can estimate a program behaviour and generate a string
- CNN are for translation invariance

Compactness: Representations are (sparse and) distributed.

- Essential to perception, compact storage and reasoning

Accounting for uncertainty: Neural nets can be stochastic to model distributions

Symbolic representation: realisation through sparse activations and gating mechanisms

WE WILL BRIEFLY COVER

Word embedding

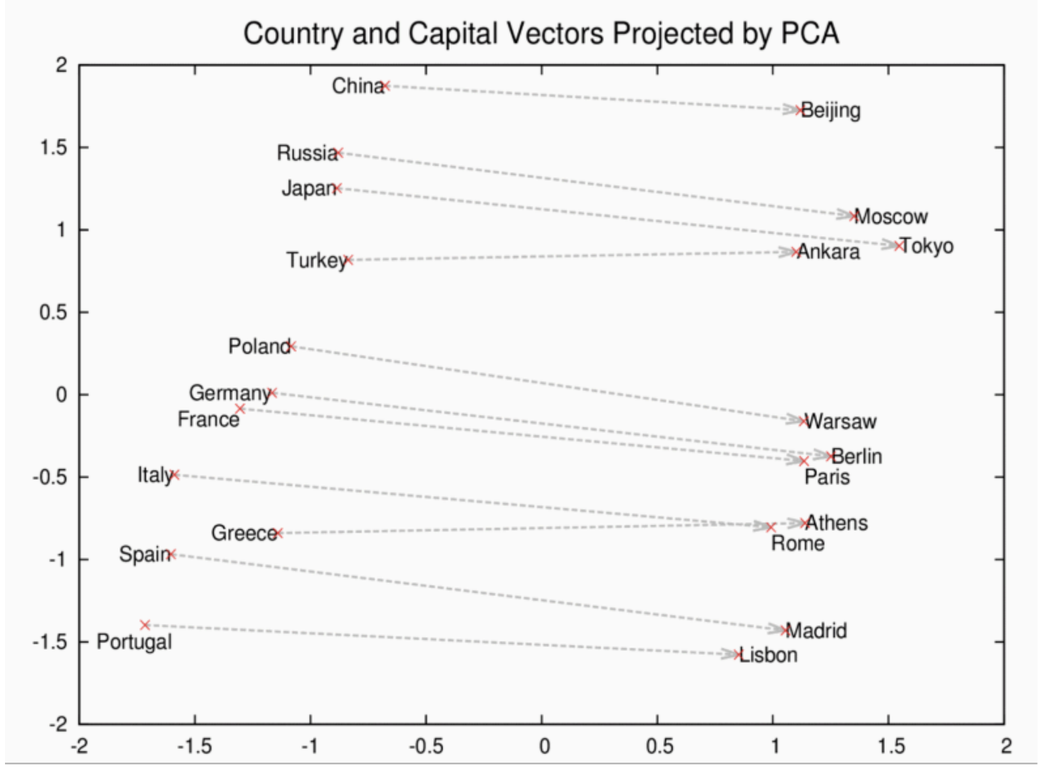
Deep autoencoder

RBM → DBN → DBM

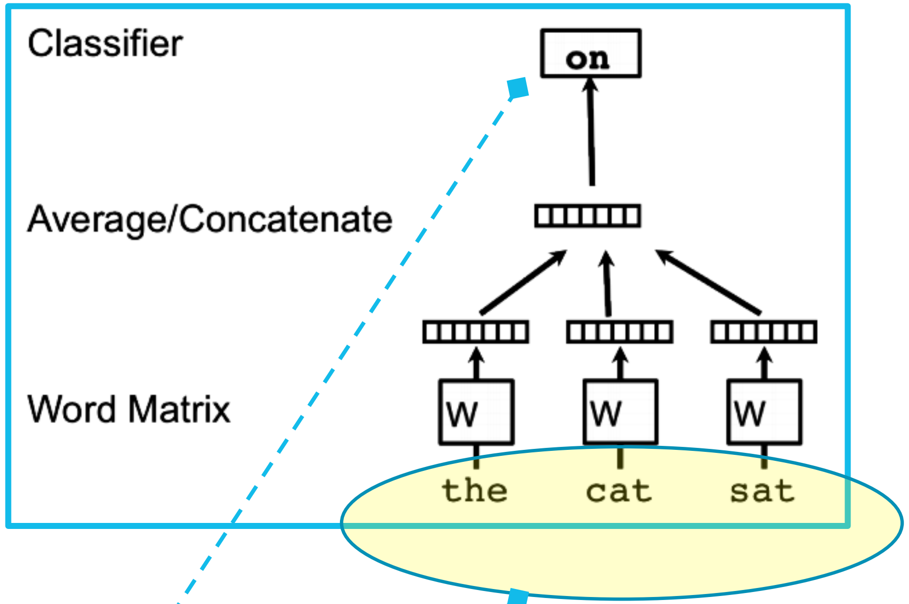
Variational AutoEncoder (VAE)

Generative Adversarial Net (GAN)

WORD EMBEDDING



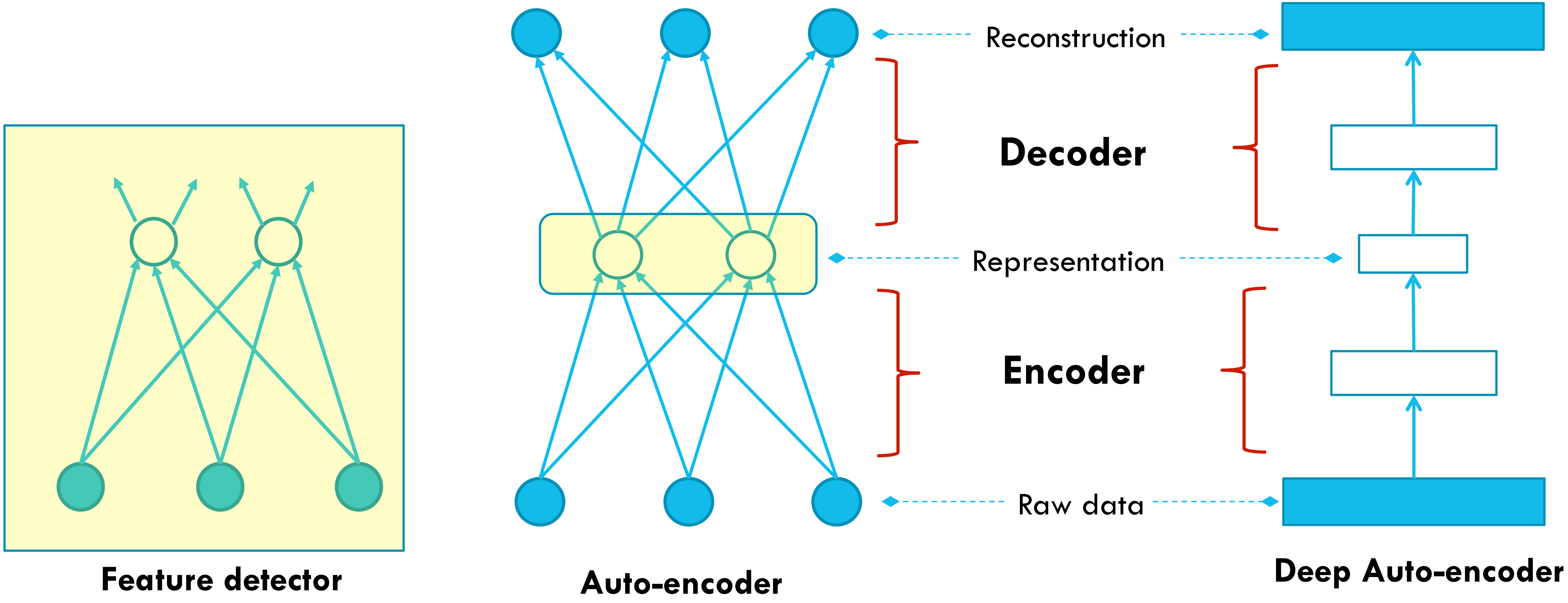
(Mikolov et al, 2013)



$$P\left(w_t \mid C_t\right) = \frac{e^{V_{w_t}^{\top} f(C_t)}}{\sum_{w \in Vocab} e^{V_w^{\top} f(C_t)}}$$

$$f(C_t) = \frac{1}{|C_t|} \sum_{w \in C_t} W_w$$

DEEP AUTOENCODER – SELF RECONSTRUCTION OF DATA



GENERATIVE MODELS

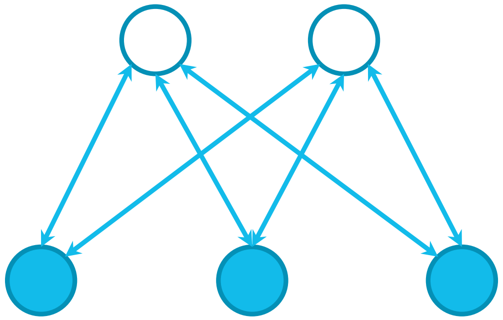
Many applications:

- Text to speech
- Simulate data that are hard to obtain/ share in real life (e.g., healthcare)
- Generate meaningful sentences conditioned on some input (foreign language, image, video)
- Semi-supervised learning
- Planning

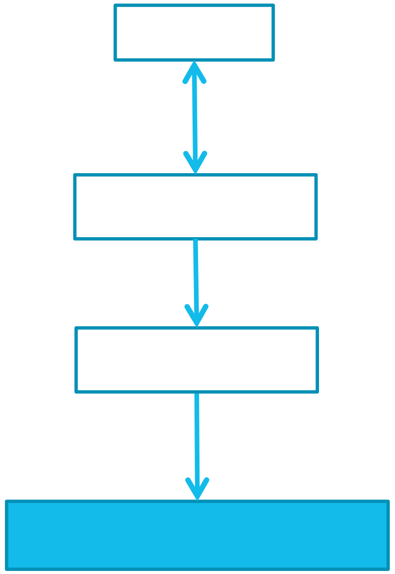
$$\mathbf{v} \sim P_{model}(\mathbf{v})$$
$$P_{model}(\mathbf{v}) \approx P_{data}(\mathbf{v})$$

A FAMILY: RBM → DBN → DBM

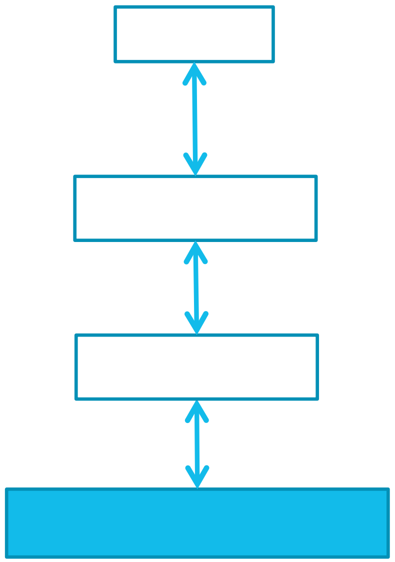
$$p\left(\mathbf{v},\mathbf{h};\psi\right) \propto \exp\left[-\underbrace{E\left(\mathbf{v},\mathbf{h};\psi\right)}_{\textit{energy}}\right]$$



Restricted Boltzmann Machine
(~1994, 2001)



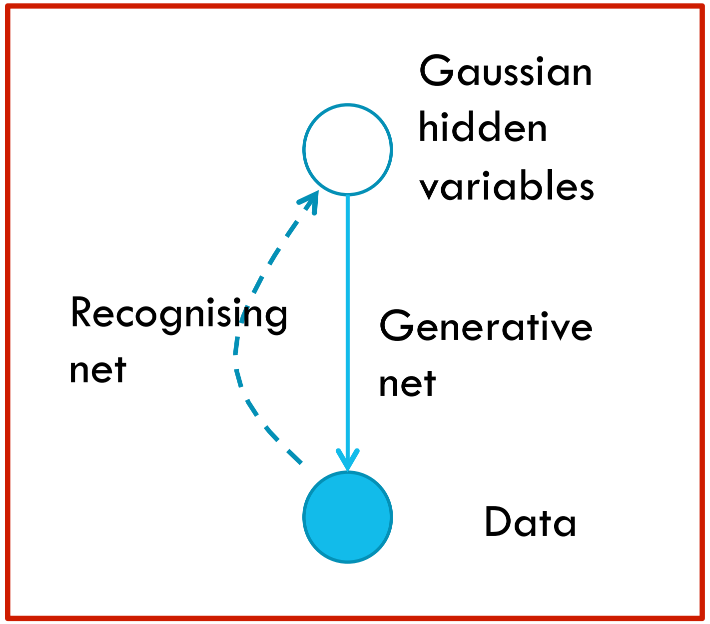
Deep Belief Net
(2006)

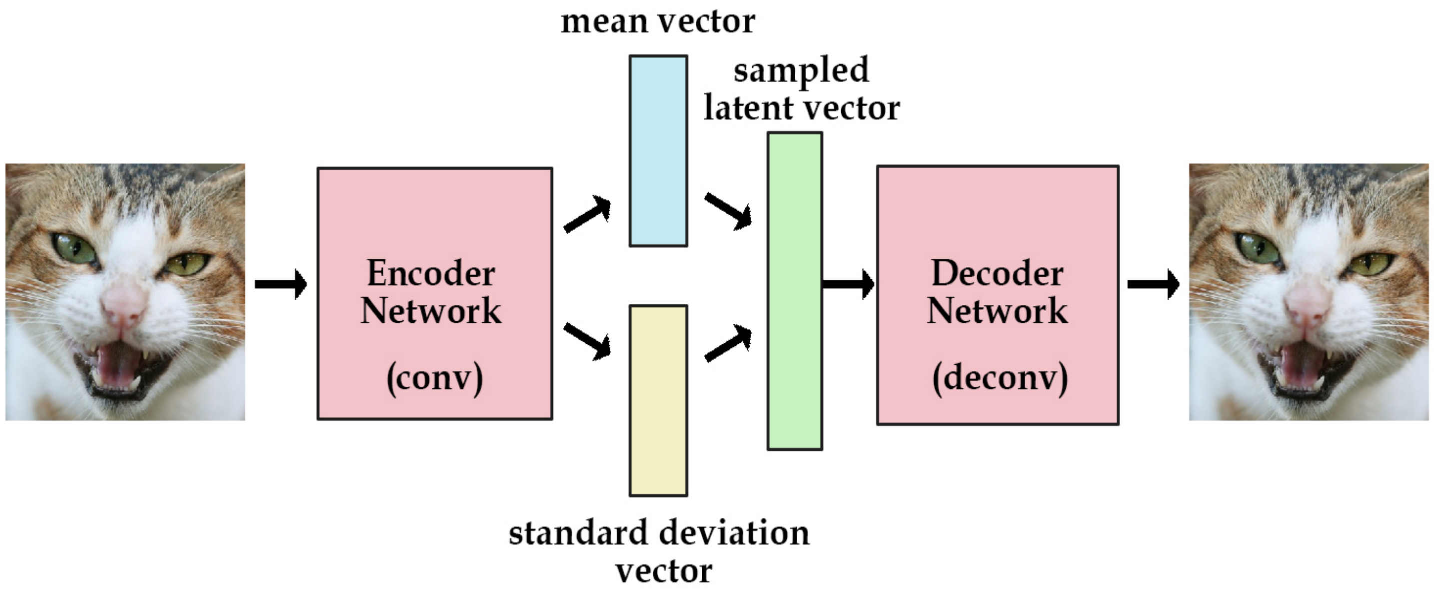


Deep Boltzmann Machine
(2009)

VARIATIONAL AUTOENCODER
(KINGMA & WELLING, 2014)

Two separate processes: generative (hidden → visible) versus recognition (visible → hidden)





http://kvfrans.com/variational-autoencoders-explained/

GAN: GENERATIVE ADVERSARIAL NETS
(GOODFELLOW ET AL, 2014)

Yann LeCun: *GAN is one of best idea in past 10 years!*

Instead of modeling the entire distribution of data, learns to map ANY random distribution into the region of data, so that ***there is no discriminator that can distinguish sampled data from real data.***

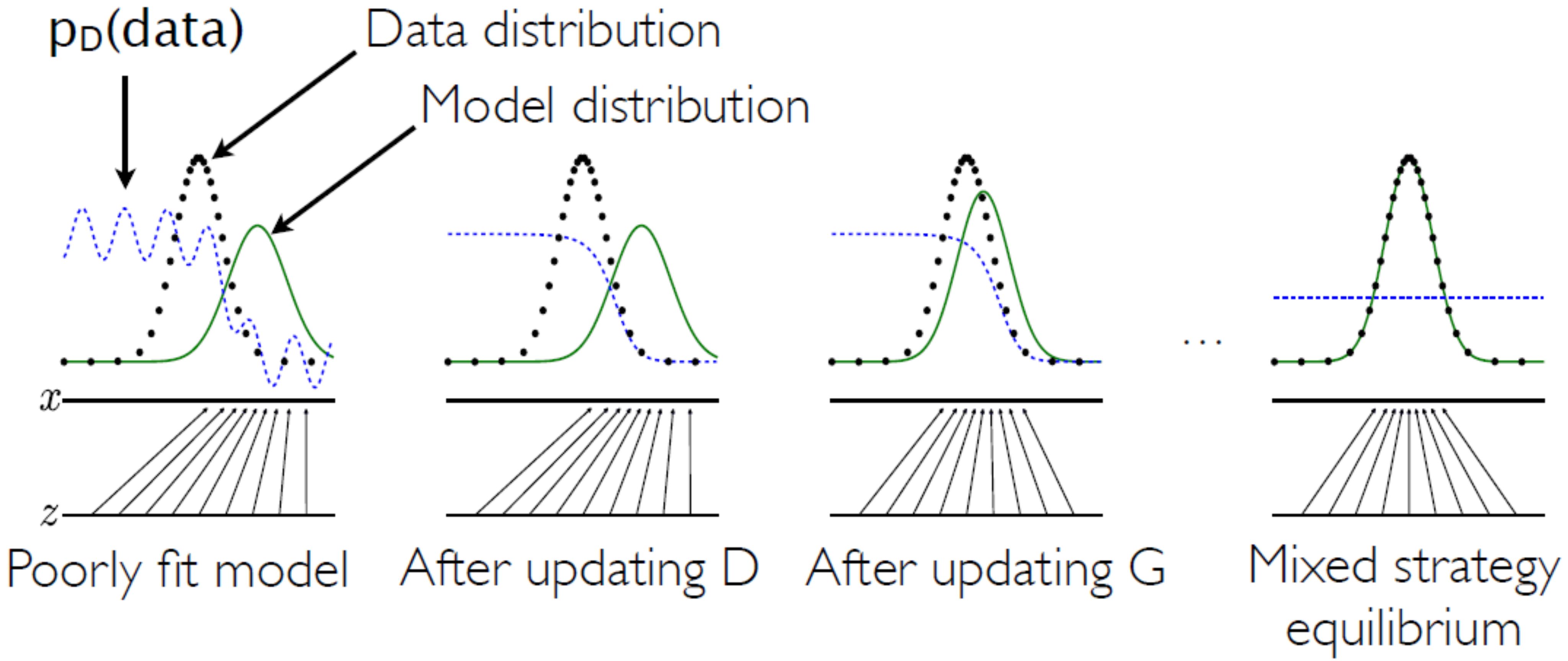
$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

Binary discriminator, usually a neural classifier

Any random distribution in any space

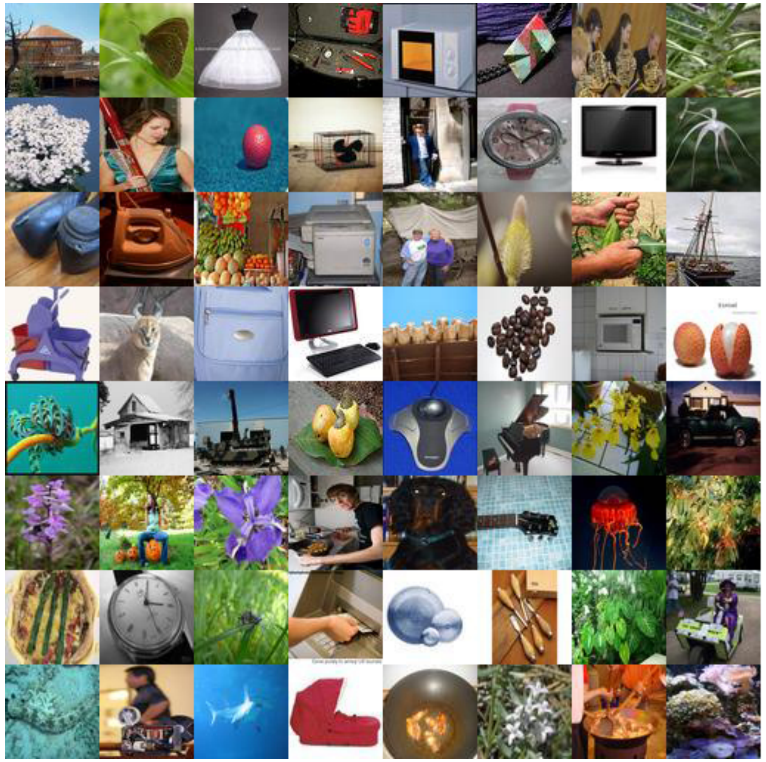
Neural net that maps z → x

GAN: LEARNING DYNAMICS
(ADAPTED FROM GOODFELLOW'S, NIPS 2014)

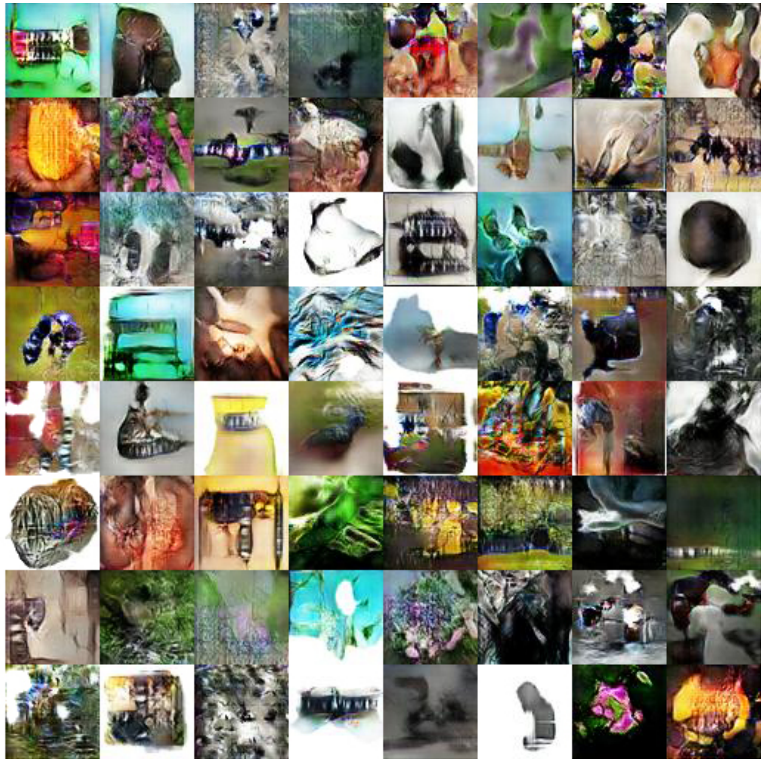


GAN: GENERATED SAMPLES

The best quality pictures generated thus far!



Real



Generated

http://kvfrans.com/generative-adversial-networks-explained/

PART III: ADVANCED TOPICS

Unsupervised learning & Generative models

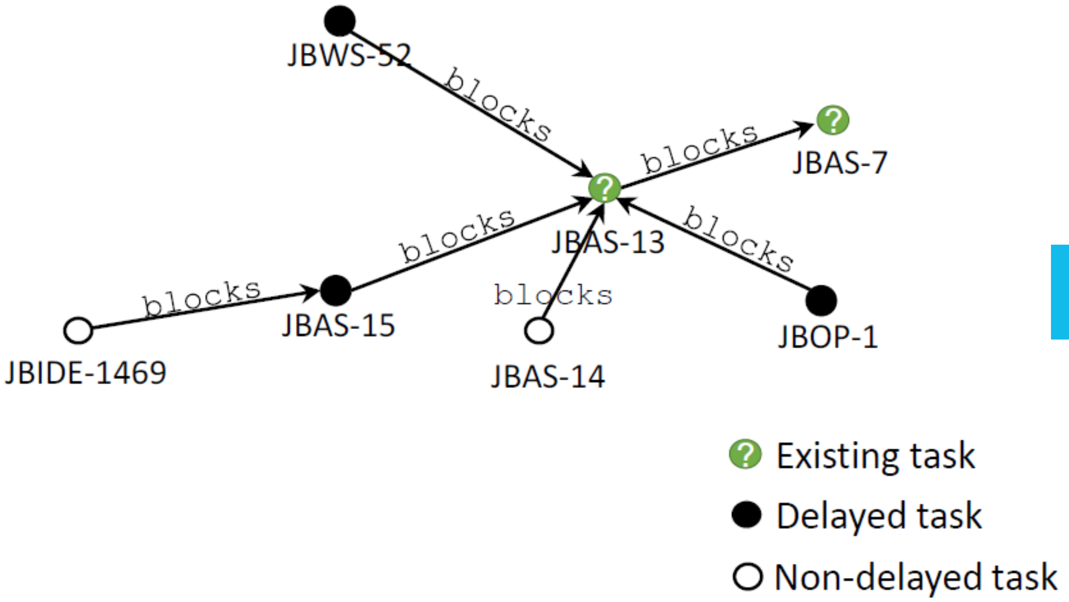
Complex domain structures: Relations (explicit & implicit), graphs & tensors

Memory, attention

How to position ourselves

EXPLICIT RELATIONS

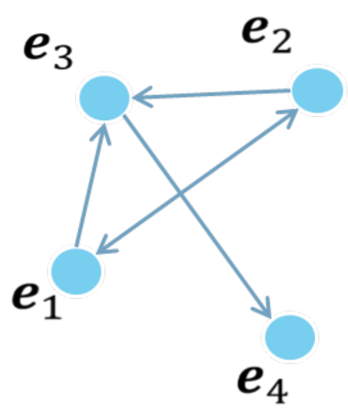
Canonical problem: **collective classification**, a.k.a. structured outputs, networked classifiers



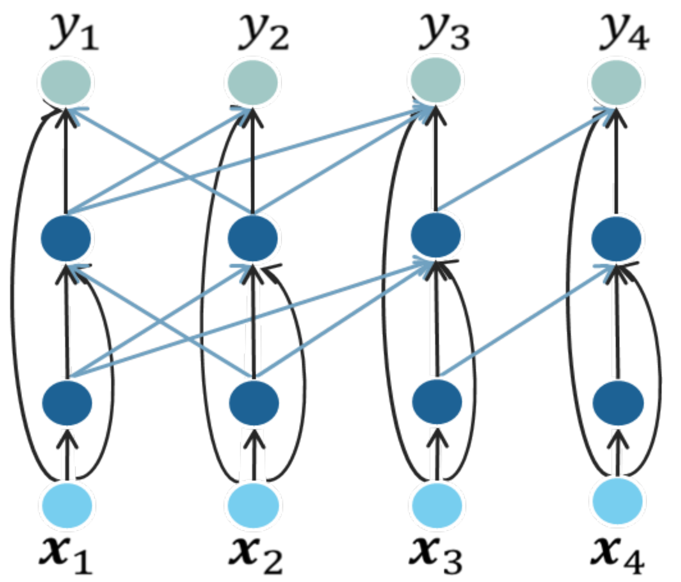
Each node has its own attributes

- Stacked inference
- Neural conditional random fields
- Column networks

STACKED INFERENCE



Relation graph

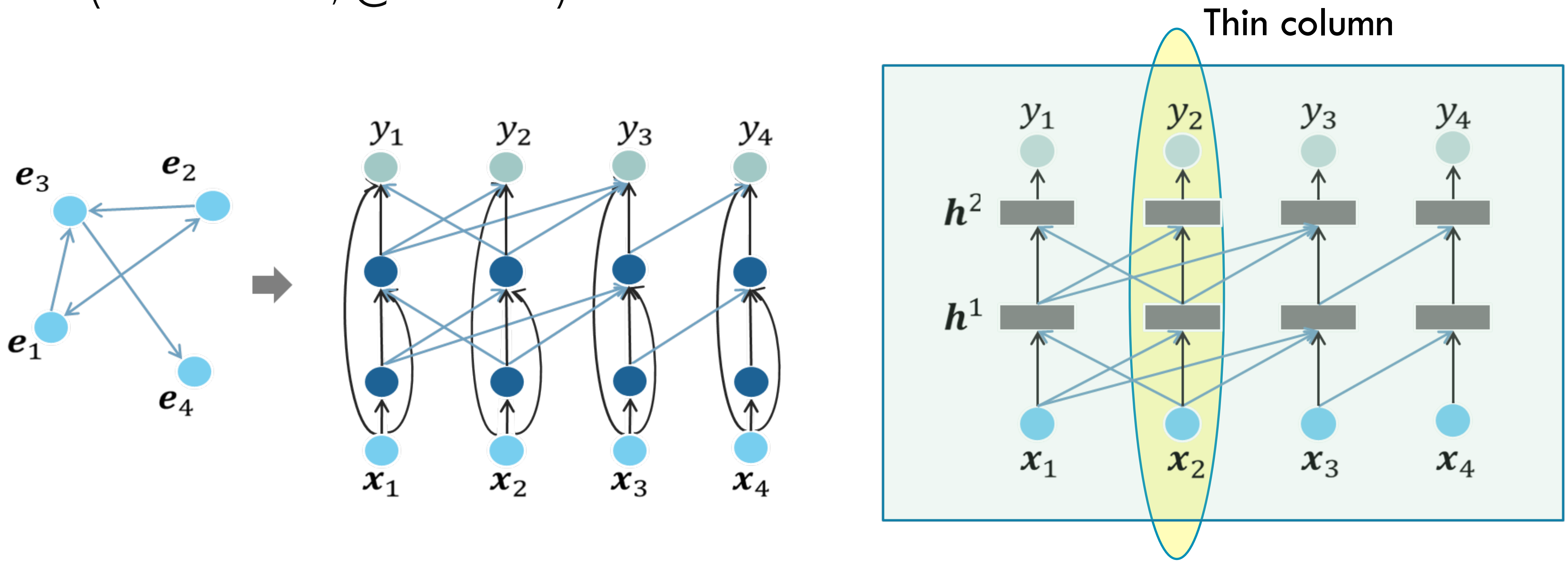


Stacked inference

Depth is achieved by stacking several classifiers.

Lower classifiers are frozen.

COLUMN NETWORKS
(PHAM ET AL, @ AAAI'16)

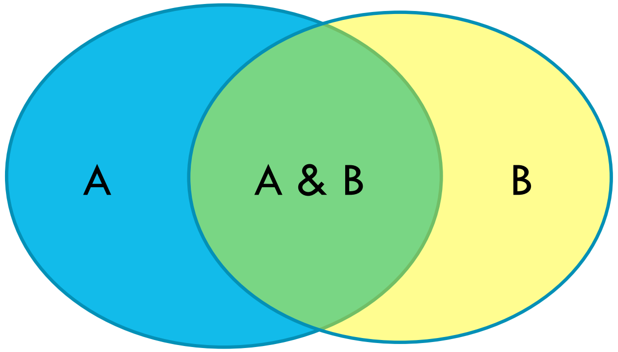


Relation graph **Stacked learning** **Column nets**

IMPLICIT RELATIONS IN CO-OCCURRENCE OF MULTI-[X] WITHIN A CONTEXT

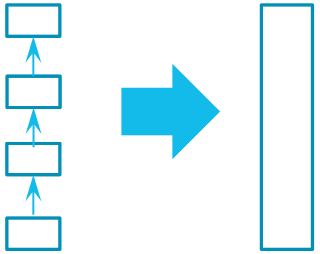
X can be:
- Labels
- Tasks
- Views/parts
- Instances
- Sources

Much of recent machine learning!

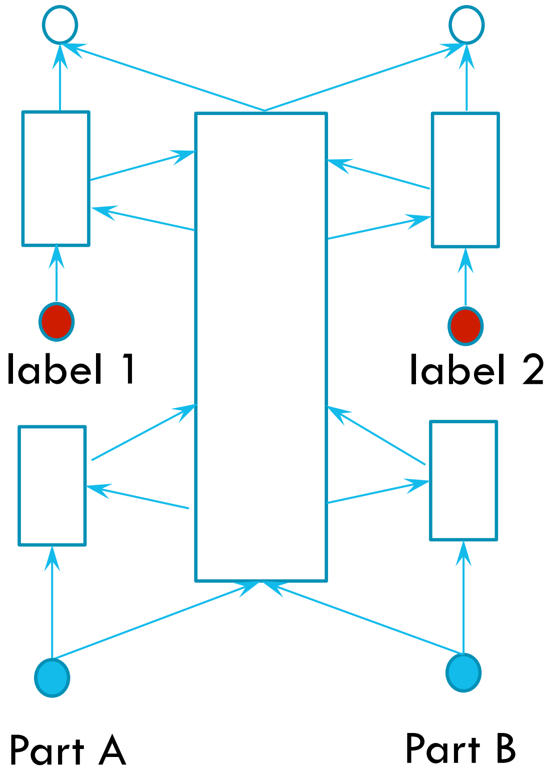


The common principle is to exploit the shared statistical strength

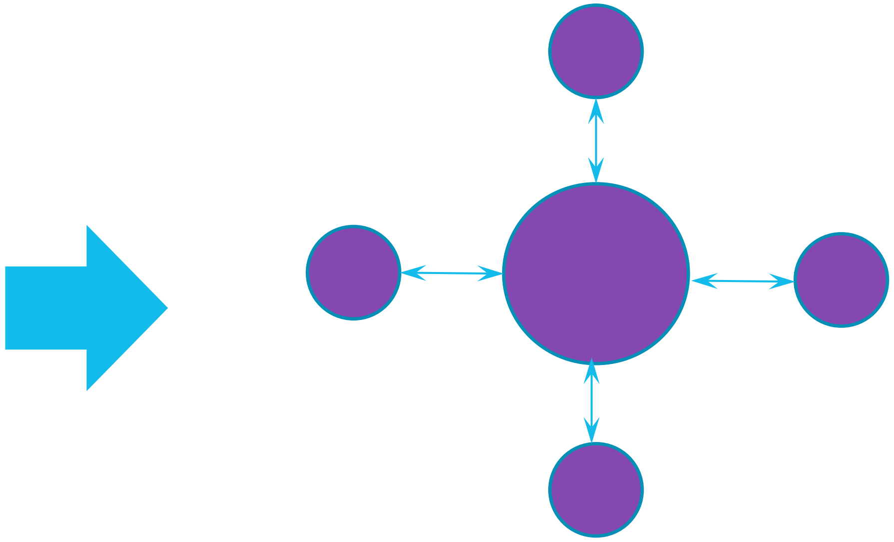
COLUMN BUNDLE FOR N-TO-M MAPPING
(PHAM ET AL, WORK IN PROGRESS)



Column



N-to-m setting



Cross-sectional star topology

GRAPHS AS DATA

Goal: representing a graph as a vector

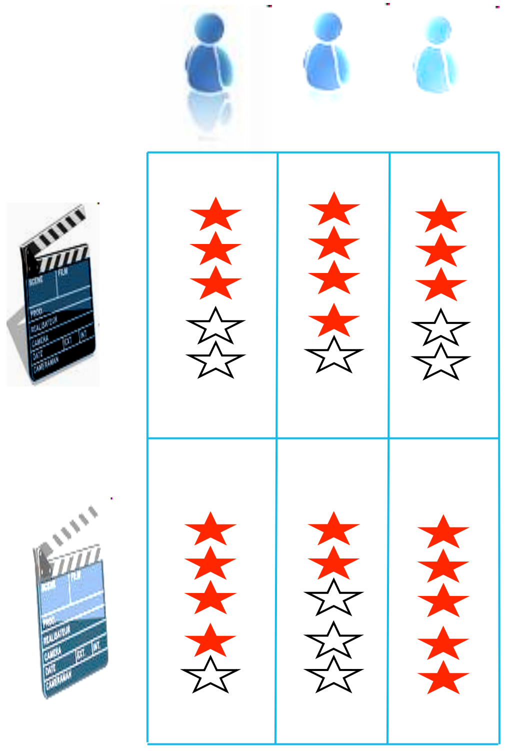
Many applications

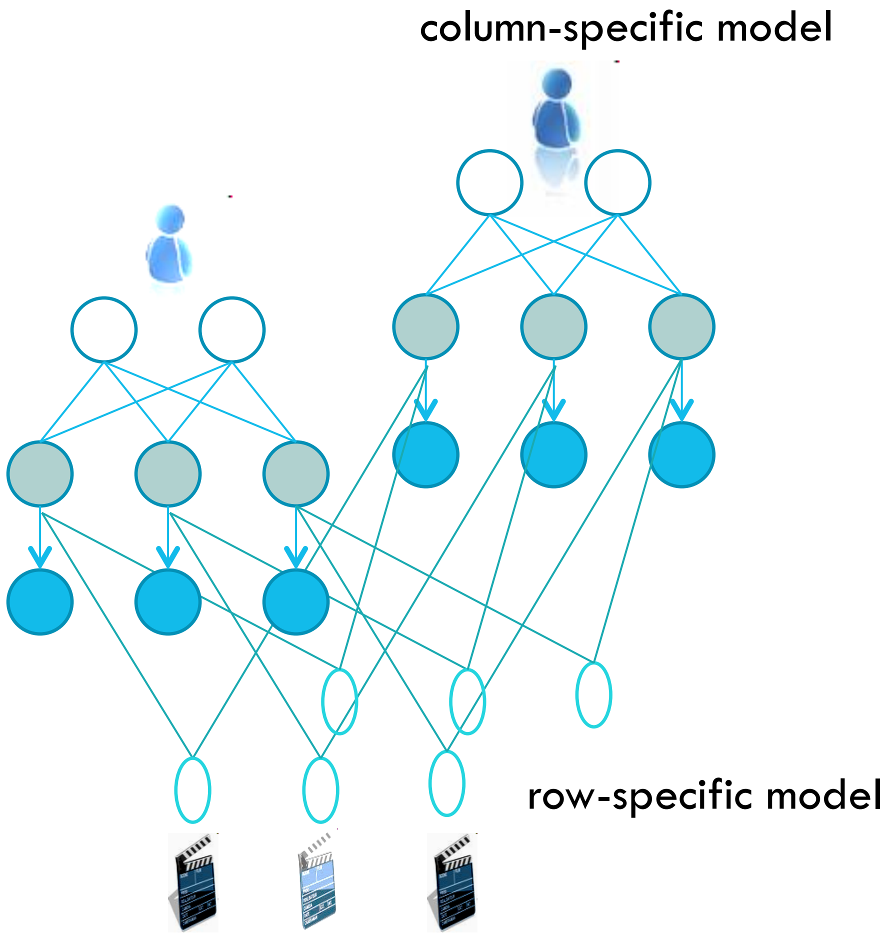
- Drug molecules
- Object sub-graph in an image
- Dependency graph in software deliverable

Recent works:

- Graph recurrent nets, similar to column nets (Pham et al, 2017).
- Graph variational autoencoder (Kipf & Welling, 2016)
- Convolutions for graph (LeCun, Welling and many others)

RBM FOR MATRIX DATA (TRAN ET AL, 2009, 2012)





PART III: ADVANCED TOPICS

Unsupervised learning & Generative models

Complex domain structures: Relations (explicit & implicit), graphs & tensors

Memory & attention

Learning to learn

How to position ourselves

ATTENTION MECHANISM

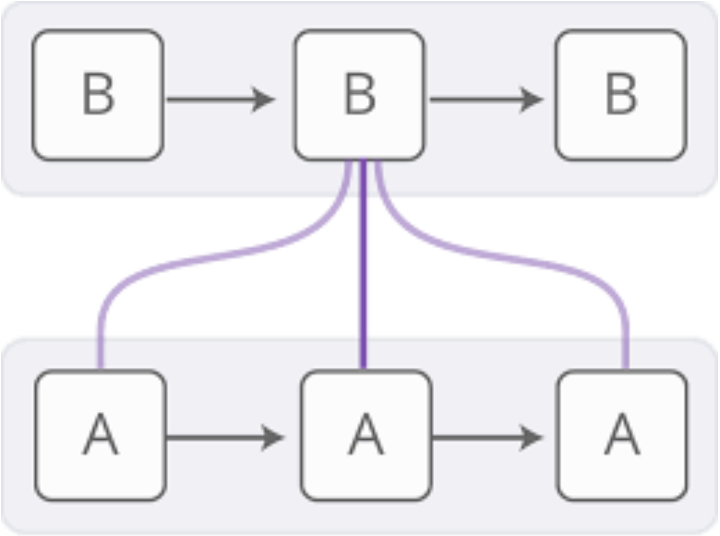
Need attention model to select or ignore certain inputs

Human exercises great attention capability – the ability to filter out unimportant noises

- Foveating & saccadic eye movement

In life, events are not linear but interleaving.

Pooling (as in CNN) is also a kind of attention



http://distill.pub/2016/augmented-rnns/

APPLICATIONS

Machine reading & question answering

- Attention to specific events/words/sentences at the reasoning stage

Machine translation

- Word alignment – attend to a few source words
- Started as early as IBM Models (1-5) in early 1990s

Speech recognition

- A word must be aligned to a segment of soundwave

Healthcare

- Diseases can be triggered by early events and take time to progress
- Illness has memory – negative impact to the body and mind

END-TO-END MEMORY NETWORKS
(SUKHBAATAR ET AL, 2015)

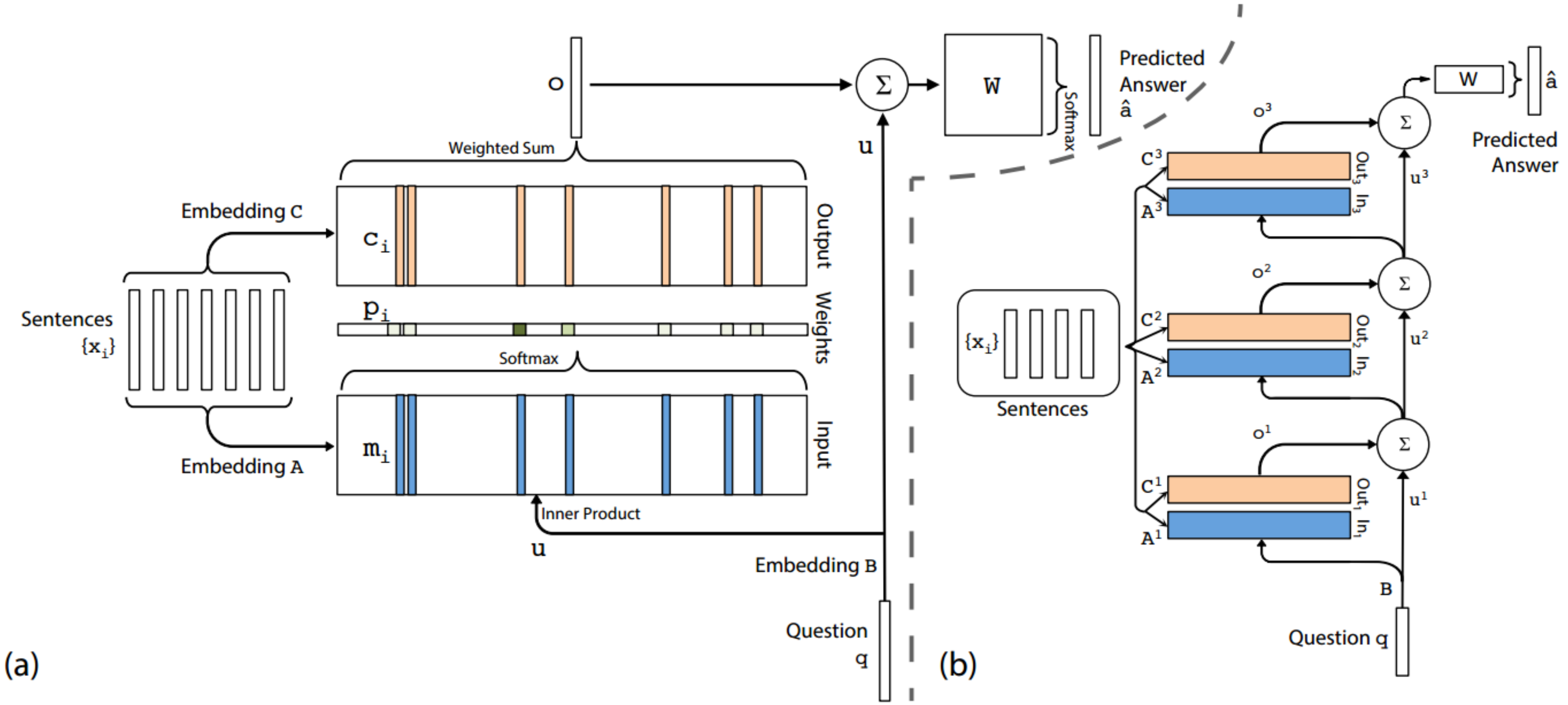


Figure 1: (a): A single layer version of our model. (b): A three layer version of our model. In practice, we can constrain several of the embedding matrices to be the same (see Section 2.2).

PART III: ADVANCED TOPICS

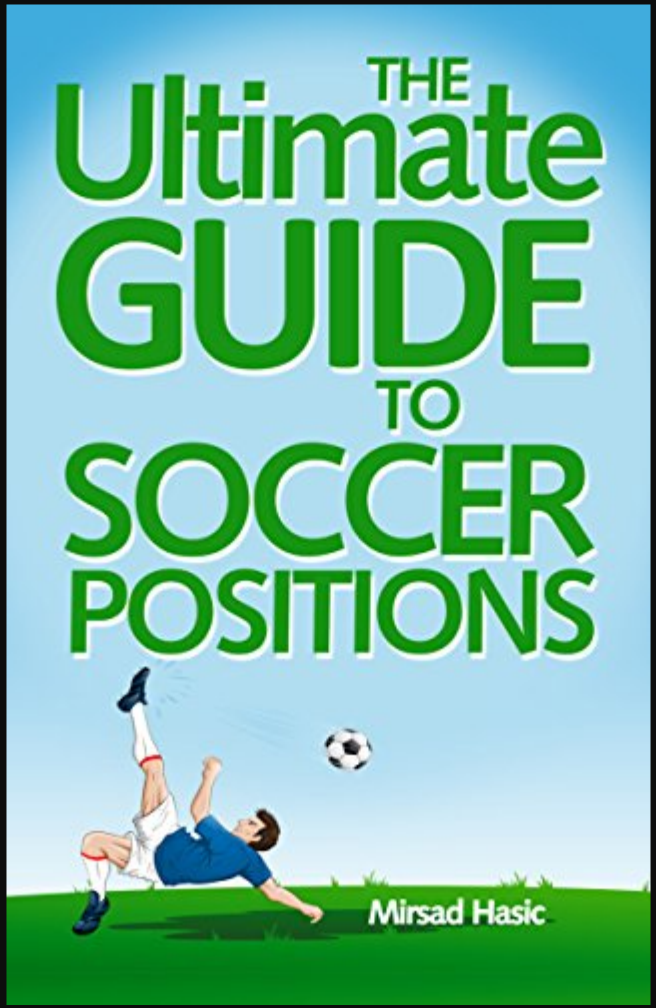
Unsupervised learning & Generative models

Complex domain structures: Relations (explicit & implicit), graphs & tensors

Memory, attention & execution

Learning to learn

How to position ourselves



POSITION YOURSELF

"[...] the dynamics of the game will evolve. In the long run, the right way of playing football is to position yourself intelligently and to wait for the ball to come to you. You'll need to run up and down a bit, either to respond to how the play is evolving or to get out of the way of the scrum when it looks like it might flatten you." (*Neil Lawrence, 7/2015, now with Amazon*)

http://inverseprobability.com/2015/07/12/Thoughts-on-ICML-2015/

THE ROOM IS WIDE OPEN

Architecture engineering

Non-cognitive apps

Unsupervised learning

Graphs

Learning while preserving privacy

Modelling of domain invariance

Better data efficiency

Multimodality

Learning under adversarial stress

Better optimization

Going Bayesian

http://smerity.com/articles/2016/architectures_are_the_new_feature_engineering.html

Thank you!

http://ahsanqawl.com/2015/10/qa/