Cross-Domain Scruffy Inference

Kenneth C. Arnold   Henry Lieberman

MIT Mind Machine Project
Software Agents Group, MIT Media Laboratory

AAAI Fall Symposium on Common Sense Knowledge
November 2010

Vision

- Informal ("scruffy") inductive reasoning over non-formalized knowledge
- Use multiple knowledge bases without tedious alignment.

We've Done This. . .

- ConceptNet and WordNet (Havasi et al. 2009)
- Topics and Opinions in Text (Speer et al. 2010)
- Code and Descriptions of Purpose (Arnold and Lieberman 2010)

but how does it work?

This Talk

- Background
- Blending is Collective Matrix Factorization.
- Singular vectors rotate.
- Other blending layouts work too.

Matrix Representations of Knowledge

	cat	dog	airplane	toaster
...IsA pet	+6	+5		
...AtLocation home	+8	+2		+1
...CapableOf fly	−3	−5	+9	
...MadeOf metal			+1	+1
fur PartOf ...	+6	+5		

Factored Inference

Filling in missing values is inference.

	cat	dog	airplane	toaster
... IsA pet	+6	+5		
... AtLocation home	+8	+2		+1
... CapableOf fly	−3	−5	+9	
... MadeOf metal			+1	+1
fur PartOf ...	+6	+5		

Factored Inference

- Represent each concept i and each feature j by k-dimensional vectors $\vec{c}_i$ and $\vec{f}_j$ such that when $A(i, j)$ is known,

$$A(i, j) \approx \vec{c}_i \cdot \vec{f}_j.$$

- If $A(i, j)$ is unknown, infer $\vec{c}_i \cdot \vec{f}_j$.
- Equivalently, stack each $\vec{c}_i$ in rows of C, same for F, then

$$A \approx CF^T.$$

Quantifying factorization quality

- Quantify the "$\approx$" in $A \approx CF^T$ as a divergence:

$$D(XY^T|A)$$

- Minimizing loss ensures that the factorization fits the data
- Many functions possible, e.g., SVD minimizes squared error:

$$D_{x^2}(\hat{A}|A) = \sum_{ij}(a_{ij} - \hat{a}_{ij})^2.$$

Collective Matrix Factorization

An analogy. . .

- Let people p rate restaurants r, represented by positive or negative values in $\|p\| \times \|r\|$ matrix A.
- Restaurants also have characteristics c (e.g., “serves vegetarian food”, “takes reservations”, etc.), represented by matrix B.
- Incorporating characteristics may improve rating prediction.
- Use the same restaurant vector to factor preferences and characteristics:

$$A \approx PR^T \qquad B \approx RC^T$$

Collective Matrix Factorization

$$A \approx PR^T \qquad B \approx RC^T$$

(A is person by restaurant, B is restaurant by characteristics)

- Collective Matrix Factorization (Singh and Gordon 2008) gives a framework for solving this type of problem
- Spread out the approximation loss:

$$\alpha D(PR^T|A) + (1-\alpha)D(RC^T|B)$$

- At $\alpha = 1$, factors as if characteristics were just patterns of ratings. At $\alpha = 0$, factors as if only qualities, not individual restaurants, mattered for ratings.

Blending is a CMF

$$A \approx PR^T \qquad B \approx RC^T$$

(A is person by restaurant, B is restaurant by characteristics)

- Can also solve with Blending:

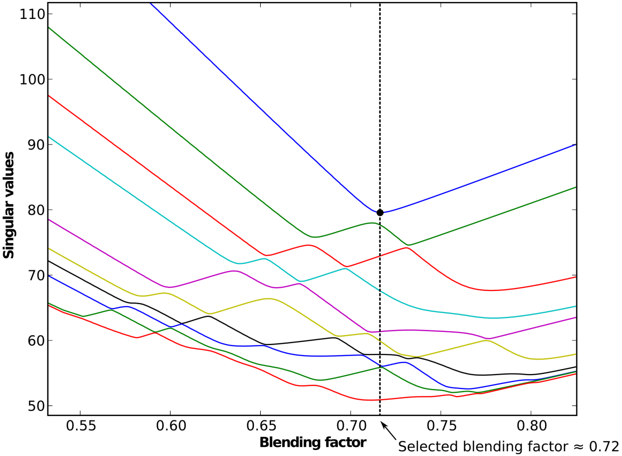
$$Z = \begin{bmatrix} \alpha A^T \\ (1-\alpha)B \end{bmatrix} \approx R \begin{bmatrix} P \\ C \end{bmatrix}^T$$

- If decomposition is SVD, loss is seperable by component:

$$D\left(R \begin{bmatrix} P \\ C \end{bmatrix}^T |Z\right) = D(RP^T|\alpha A^T) + D(RC^T|(1-\alpha)B)$$

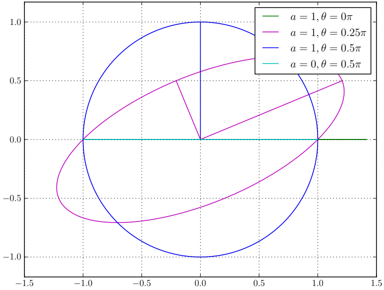
$\Rightarrow$ Blending is a kind of Collective Matrix Factorization

Veering


Singular values
110
100
90
80
70
60
50
0.55
0.60
0.65
0.70
0.75
0.80
Blending factor
Selected blending factor ≈ 0.72

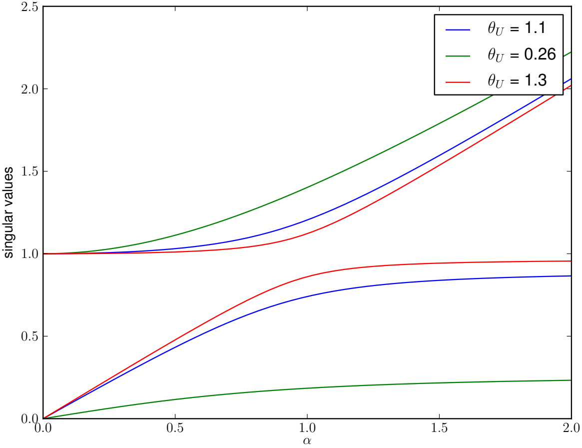
Blended Data Rotates the Factorization

- What happens at an intersection point?
- Consider you're blending X and Y. Start with $X \approx AB^T$; what happens as you add in Y?
- First add in the new space that only Y covered.
- Now data is off-axis, so rotate the axes to align with the data.

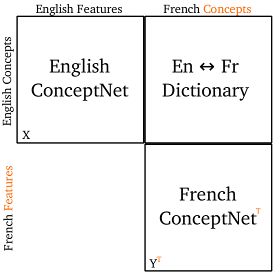


Veering

"Veering" is caused by singular vectors of the blend rotating between corresponding singular vectors of the source matrices.



Bridge Blending



General bridge blend:

$$\begin{bmatrix} X & Y \\ & Z \end{bmatrix} \approx \begin{bmatrix} U_{XY} \\ U_{0Z} \end{bmatrix} \begin{bmatrix} V_{X0} \\ V_{YZ} \end{bmatrix}^T = \begin{bmatrix} U_{XY} V_{X0}^T & U_{XY} V_{YZ}^T \\ U_{0Z} V_{X0}^T & U_{0Z} V_{YZ}^T \end{bmatrix}$$

Again, loss factors:

$$\begin{aligned} D(\hat{A}|A) = & D(U_{XY} V_{X0}^T | X) + D(U_{XY} V_{YZ}^T | Y) + \\ & D(U_{0Z} V_{X0}^T | 0) + D(U_{0Z} V_{YZ}^T | Z) \end{aligned}$$

V_{YZ} ties factorization of X and Z together through bridge data Y.
Could use weighted loss in empty corner.

Summary

- Blending is a Collective Matrix Factorization
- "Veering" indicates singular vectors rotating between datasets

- What's next?
  - CMF permits many objective functions, even different ones for different input data. What's appropriate for commonsense inference?
  - Incremental?
  - Can CMF do things we thought we needed 3rd-order for?