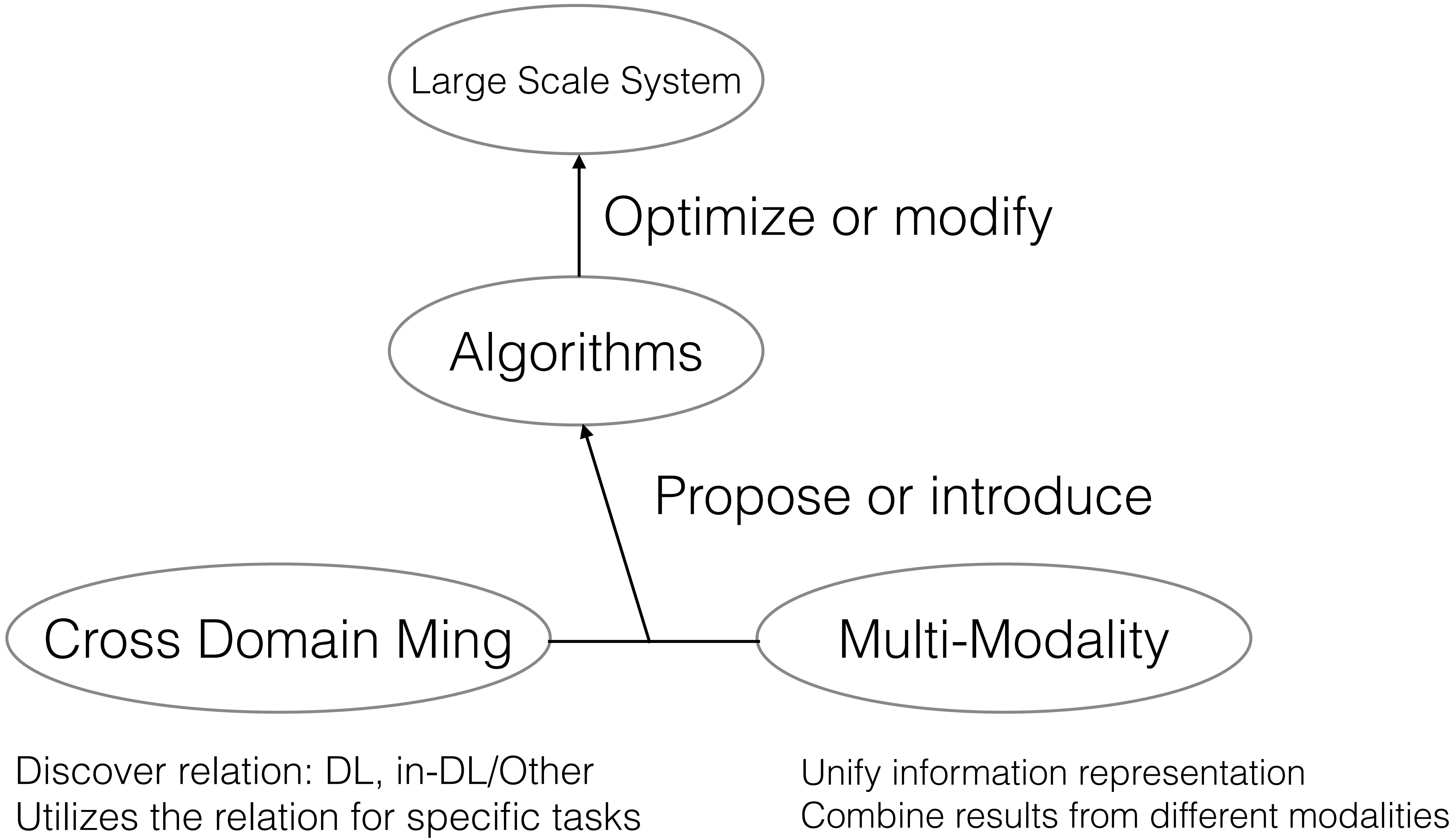

Large Scale System
Optimize or modify
Algorithms
Propose or introduce
Cross Domain Ming
Multi-Modality
Discover relation: DL, in-DL/Other
Utilizes the relation for specific tasks
Unify information representation
Combine results from different modalities

Multimodality

- Unify/convert materials in different modalities into a uniformed encoding.

  Convert image into texts
  Convert images and text into trees.
  Translation of different languages
  …

- Combine results from different modality

  Get activity from images and category of the text and combine the two to predict its popularity

Multimodality

Uniform encoding:

- There are some work that turn a picture into a sentence. We may try that on Pinterest data.
- Try to describe a picture with more sophisticated methods — objects in it, colorful or not, in-door or not, activities, etc.

Combined result:

- Classify pictures according to Pinterest category
- Combine the picture classification result with the text classification results.

Schedule:

- Test combined result first and see whether it work.
- Try to build converting algorithms.
- Test different uniform encoding later.

Cross Domain Ming

- Direct relations
  Two domains share common data points — eBay-Amazon, Instagram-Facebook
  We have some extra information compared to single domain.
  Details
- Indirect relations
  Two domains that share same feature — forums and blogs, twitter and Facebook
  We may treat the data with the same feature as a same data point and apply both classifiers or any other operations on the data and gain the link
  Details
- Other relations
  Any other relations. Hard to formalize but have vast topics to explore.
  One example may be here.

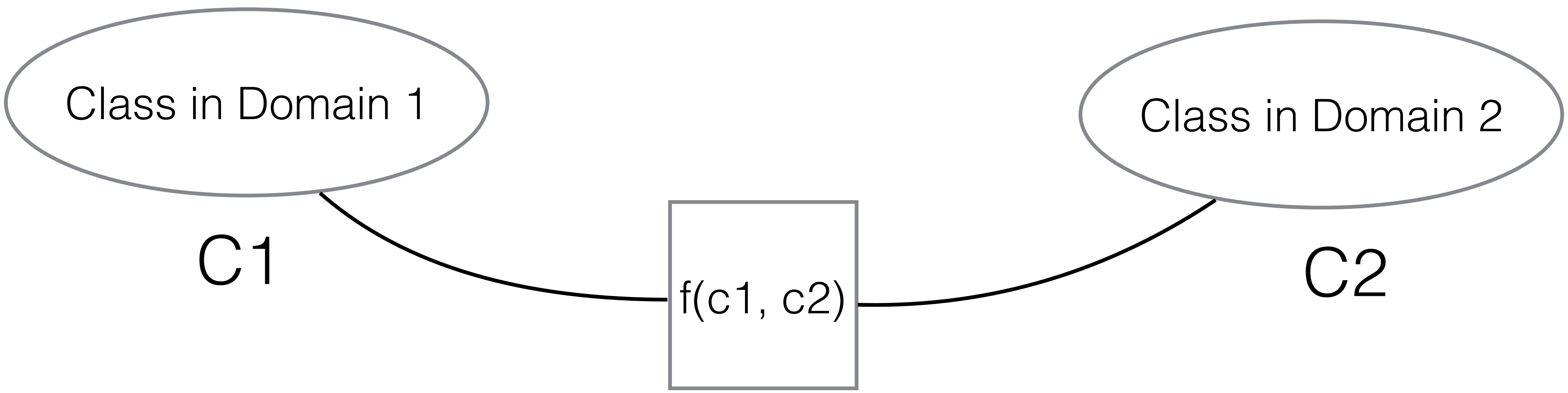
Direct/Indirect Relations

Output of classifier 1

class 1	class 2	...
probability 1	probability 2	....

Output of classifier 2

Class 1	Class 2	...
Probability 1	Probability 2	....



- Maximize P(class) when classifying domain 1
- Similar trick can be applied in indirect relations, except class 1 in the table contains all the data points that is classified to class 1 by classifier 1
- Or we may apply to EM algorithm to estimate the CPD

	class 1	class 2	...
Class 1	$\frac{class1 \cap Class1}{class1 \cup Class1}$	$\frac{class2 \cap Class1}{class2 \cup Class1}$	$\dots$
Class 2	$\frac{class1 \cap Class2}{class1 \cup Class2}$	$\frac{class2 \cap Class2}{class2 \cup Class2}$	$\dots$
...	$\dots$	$\dots$	$\dots$

Direct/Indirect Relations

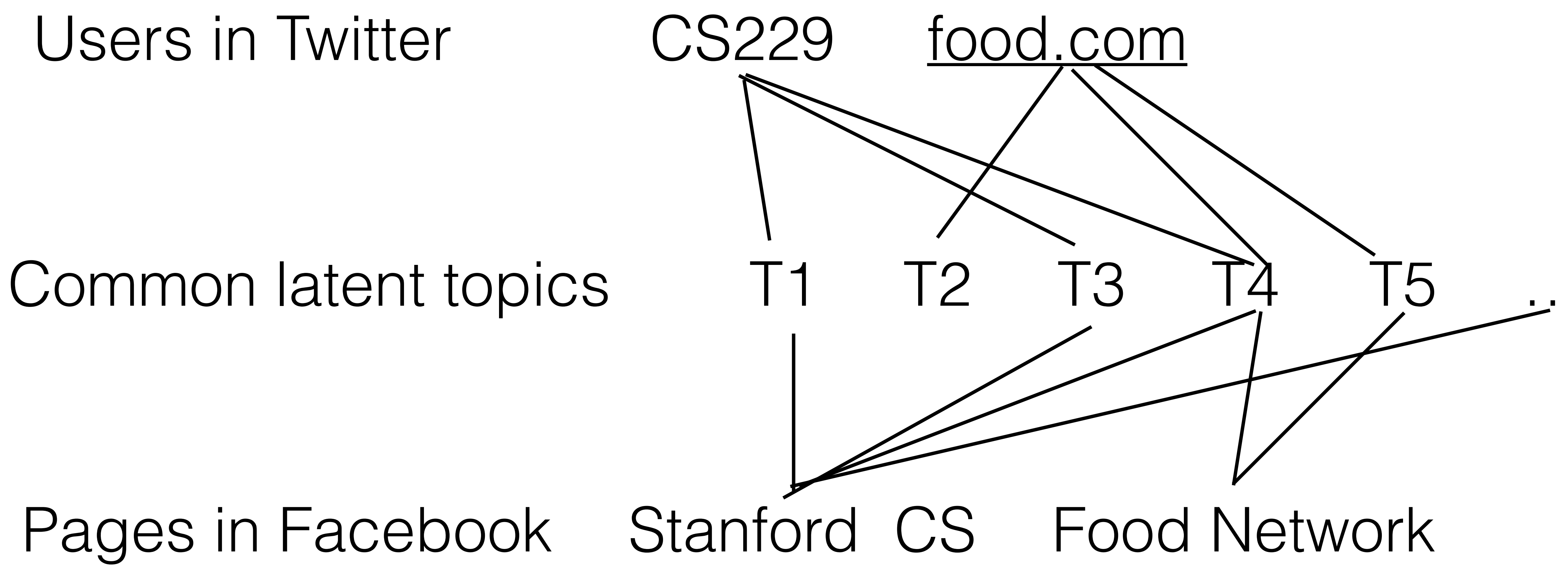
Scaling Issue:

- The CPD table increases fast when the number of classes increases. (n^mif there are both n classes in m domains)
- To reduce the computational cost, we only consider top k classes in the CPD. (The cost reduces to k^m.)

Schedule:

- Construct a data set where two domains are directly related and their classes are highly related.
- Test the proposal. Hope it will work.
- If it works, find more data set and test the how result is affected by the class relations.
- Expand to indirectly related data sets. See whether EM is needed and whether EM will help.

Other Relations



- Try to find a common set of latent topics that composes topics in two domains.
- The relation between the two domain is the latent common topics.
- For a new user in twitter who links his Facebook account, we can recommend users in twitter that have similar composition with the pages he followed in Facebook.
- But, how do we match the topics? Is any direct/indirect link a must?

Questions

- What data sources do we have?
- How are they formatted originally?
- Is there any topics that are already chosen? Is the proposed ones fit?
- What are the interested tasks? More tasks, more possible solutions.
- Any other people in this field? Previous works?