IBM Research

A Graph-based Framework for Multi-Task Multi-View Learning

Jingrui He, Rick Lawrence
Machine Learning Group
IBM T.J. Watson Research Center

Motivation

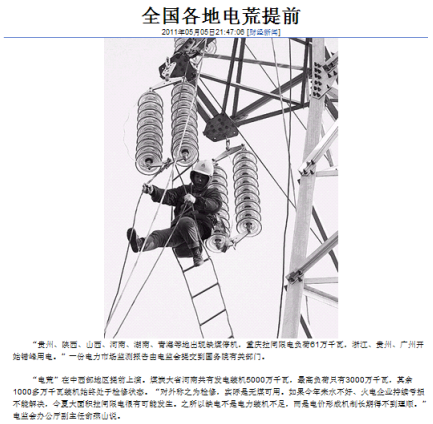
- Feature heterogeneity
  - A single task with features in multiple views
  - Multi-view learning
- Task heterogeneity
  - Multiple tasks related to each other in one way or another
  - Multi-task learning
- ***Dual heterogeneity*** in real applications
  - Feature heterogeneity & task heterogeneity
  - Shared views vs. task-specific views
  - ***Multi-Task Multi-View (M^2TV) Learning***

Main Contributions

- Problem definition
  - ***M²TV***: multiple related tasks with both shared and task-specific views
- Framework
  - ***GraM²***: a graph-based framework for *M²TV*
- Algorithm
  - ***IteM²***: an effective algorithm for solving *GraM²*

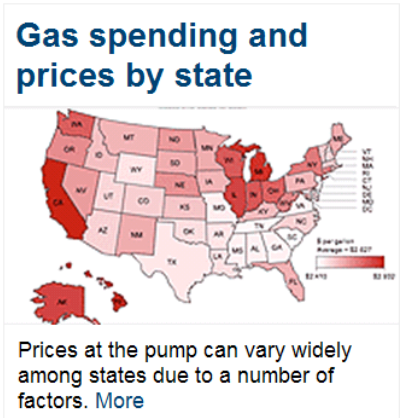
Example 1: Image Classification

Task 1: classification of images on Chinese web sites



Common view: image features e.g., color histogram

Task 2: classification of images on English web sites



Multi-task Multi-view Learning

Chinese tokens

Images

Task 1 specific view: TF-IDF representation of Chinese text content

English tokens

Images

Task 2 specific view: TF-IDF representation of English text content

Example 2: Music Classification

Task 1: classification of Chinese songs

[ti:爱的供养]
[al:宫锁心玉主题曲]
歌曲:爱的供养
歌手:杨幂
作词:于正作曲:谭晓
宫锁心玉主题曲
把你捧在手上虔诚地焚香
剪下一段烛光将经纶点亮
不求落气回肠只求爱一场
爱到最后受了伤哭得好绝望
我用尽一生一世来将你供养
只期盼你停住流转的目光
请赐予我无限爱与被爱的力量
让我能安心在菩提下静静的观想
music~
把你放在心上合起了手掌
默默乞求上苍指引我方向
不求地久天长只求在身旁
累了醉倒温柔乡轻轻地梵唱
人世间有太多的烦恼要忘
苦海中飘荡看你那旧时的模样
一回头发现早已踏出了红尘万丈

Common view: audio features e.g., pitch-class profile

Task 2: classification of English songs

The wheels on the bus go round and round,
round and round,
round and round.
The wheels on the bus go round and round,
all through the town.

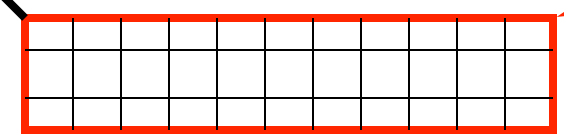
The wipers on the bus go Swish, swish, swish;
Swish, swish, swish;
Swish, swish, swish.
The wipers on the bus go Swish, swish, swish,
all through the town.

The horn on the bus goes Beep, beep, beep;
Beep, beep, beep;
Beep, beep, beep.
The horn on the bus goes Beep, beep, beep,
all through the town..

Multi-task Multi-view Learning

Chinese tokens

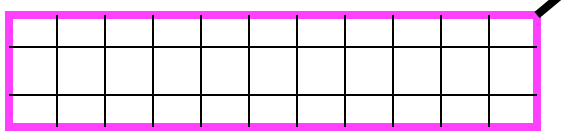
Songs



Task 1 specific view: TF-IDF representation of Chinese song lyric

English tokens

Songs



Task 2 specific view: TF-IDF representation of English song lyric

Example 3: Document Classification

Task 1: classification of **Tweets**

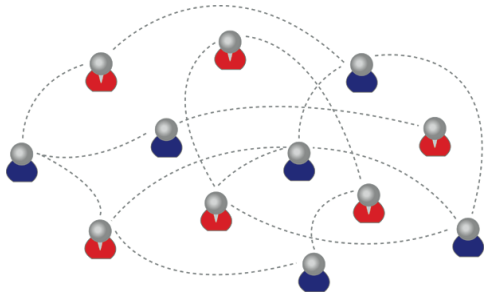
Ben Tepfer: 8:14 PM Nov 14th via Twitter for iPhone from Eastsid
Syracuse
URGENT: I just bit into a scone from @starbucks to find over 10 staples baked into it. Please RT and be careful.

Common view: TF-IDF representation

Task 2: classification of **Conventional Documents**

RAISE your hand if you remember when Starbucks seemed cool. Anyone?
Think back. To before the planet groaned with 17,000 Starbucks shops. Before the pumpkin spice lattes and the Ciao Amore CDs.

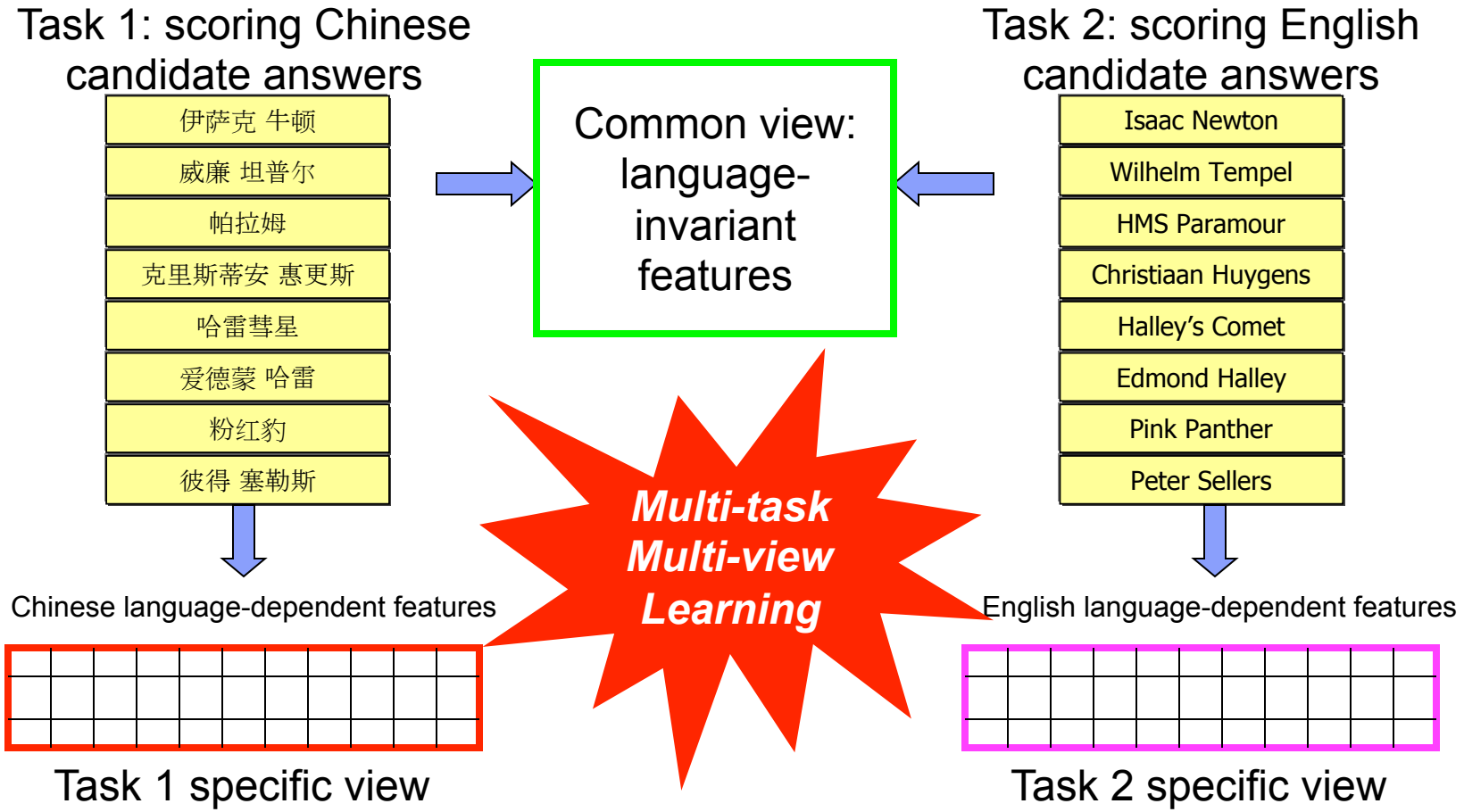
e.g. NY Times



Task 1 specific view: social network for the Twitter user

Multi-task Multi-view Learning

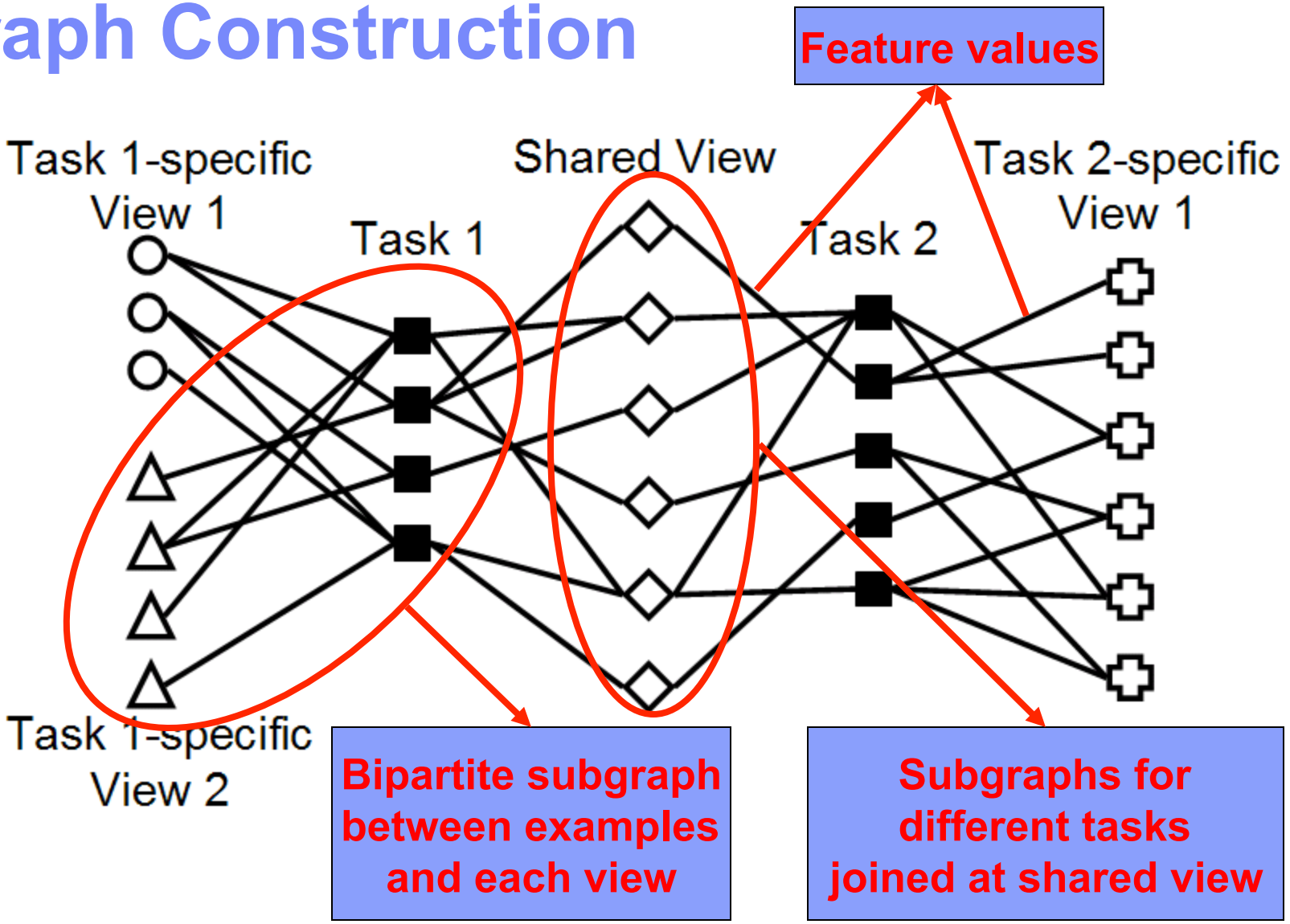
Example 4: Watson Candidate-Answer Scoring



Roadmap

- Motivation & examples
- *GraM*^{2}: a graph-based framework
- *IteM*^{2}: the proposed algorithm
- Experimental results
- Conclusion

Graph Construction



Objective Function

- Two types of functions
  - $g_i(\cdot)$: functions defined on the examples
  - $f_{ik}(\cdot)$: functions defined on the features
- Three types of constraints on the graph
  - Label consistency on the graph
    - Between $g_i(\cdot)$ and $f_{ik}(\cdot)$
  - Consistency with the label information
    - Between $g_i(\cdot)$ and the known class labels
  - Similarity on the common views
    - Between $f_{ik}(\cdot)$ of shared views for different tasks

Interpretation of *GraM²*

$$Q(f,g) = \sum_{i=1}^{T} C_i + b \sum_{i=1}^{T} \sum_{j=1}^{T} \sum_{k \in S_{ij}} \|f_{ik} - f_{jk}\|^2$$

Similarity on the common views

$$C_i = \sum_{k=1}^{V_i} a_{ik} C_{ik} + \mu_i \|g_i - y_i\|^2$$

Label consistency on the graph

Consistency with known label info

$$h^* = [h_1^*, h_2^*]^T = \arg\min_{h \in \mathcal{H}} \|h\|_{\mathcal{H}}^2 + \sum_{i=1}^{2} \mu_i \sum_{l=1}^{m_l} (h_i(x_{il}) - y_{il})^2$$

Functions defined on BOTH examples and features

Roadmap

- Motivation & examples
- *GraM*^{2}: a graph-based framework
- ***IteM*^{2}: the proposed algorithm**
- Experimental results
- Conclusion

Basic Idea

- Iterative procedure for updating both $g_i(\cdot)$ and $f_{ik}(\cdot)$
  - Update $f_{ik}(\cdot)$ for non-shared views
  - Update $f_{ik}(\cdot)$ for shared views
  - Update $g_i(\cdot)$ based on $f_{ik}(\cdot)$ of all the views
- Assign class labels based on the sign of $g_i(\cdot)$

Updating Shared Views

Indices of tasks sharing the k^{th} view with the i^{th} task

$$\frac{\partial Q(f,g)}{\partial f_{ik}} = 2a_{ik}f_{ik} - 2a_{ik}L_{ik}^{T}g_i + 4b\sum_{s\in I_{ik}}(f_{ik} - f_{sk})$$

Jointly update tasks sharing the k^{th} view

$$A_3 = A_2 A_1^{-1}$$

$$A_1^{-1} = \frac{1}{a^2 + 2rab}\begin{bmatrix} a+2b & 2b & \cdots & 2b \\ 2b & a+2b & \cdots & \vdots \\ \vdots & \vdots & \ddots & 2b \\ 2b & \cdots & 2b & a+2b \end{bmatrix}$$

Each column of A_3 corresponds $f_{ik}(\cdot)$ of tasks sharing the k^{th} view

$d_k \times |I_k|$

the i^{th} column set to $a_{I_k(i)k}L_{I_k(i)k}^{T}g_{I_k(i)}$

Properties of *IteM²*

- Optimality and convergence
  - Given sufficiently large number of iteration steps, *IteM²* will converge to the optimal solution of the objective function in *GraM²*
- Time complexity
  - $O(n_{iter}\left(\sum_{i=1}^{T} n_i \sum_{k=1}^{V_i} d_{ik} + d_S T^2 + V T^3\right) + \sum_{i=1}^{T} n_i \log n_i)$

Number of features in the k^{th} view of the i^{th} task

The total number of features in shared views

The total number of views

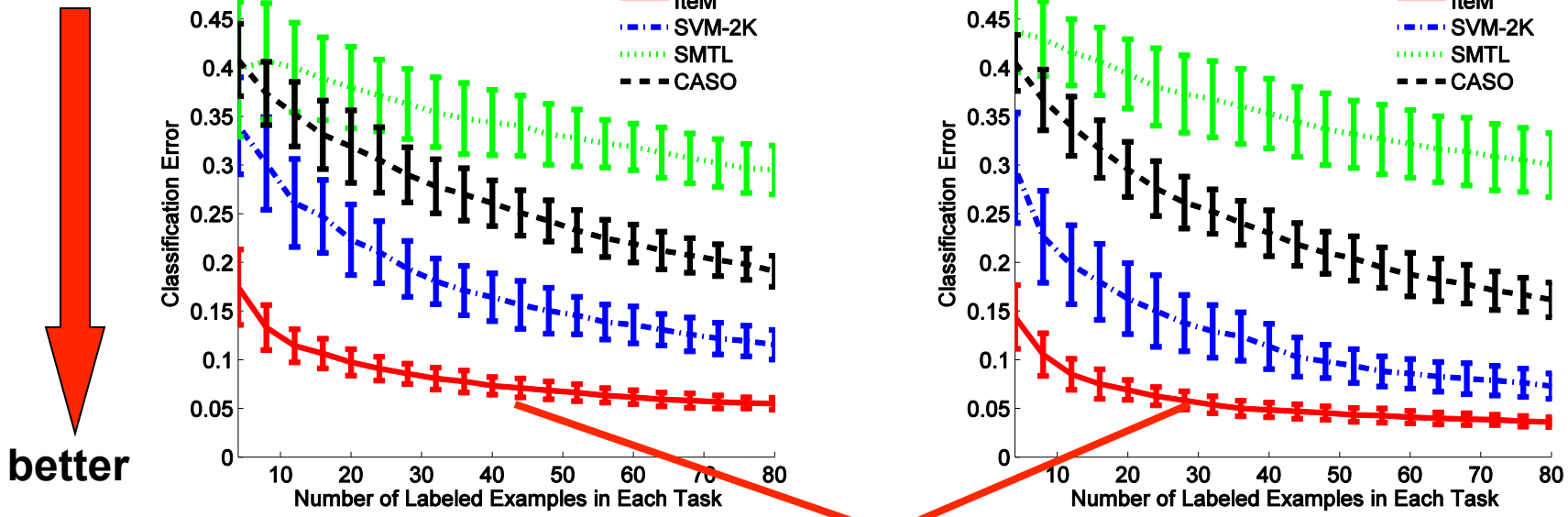
Roadmap

- Motivation & examples
- *GraM*^{2}: a graph-based framework
- *IteM*^{2}: the proposed algorithm
- **Experimental results**
- Conclusion

Two Tasks with Non-identical Views

- Data set: 20 newsgroups
- Common view: common vocabulary
- Task-specific view: task-specific vocabulary

worse

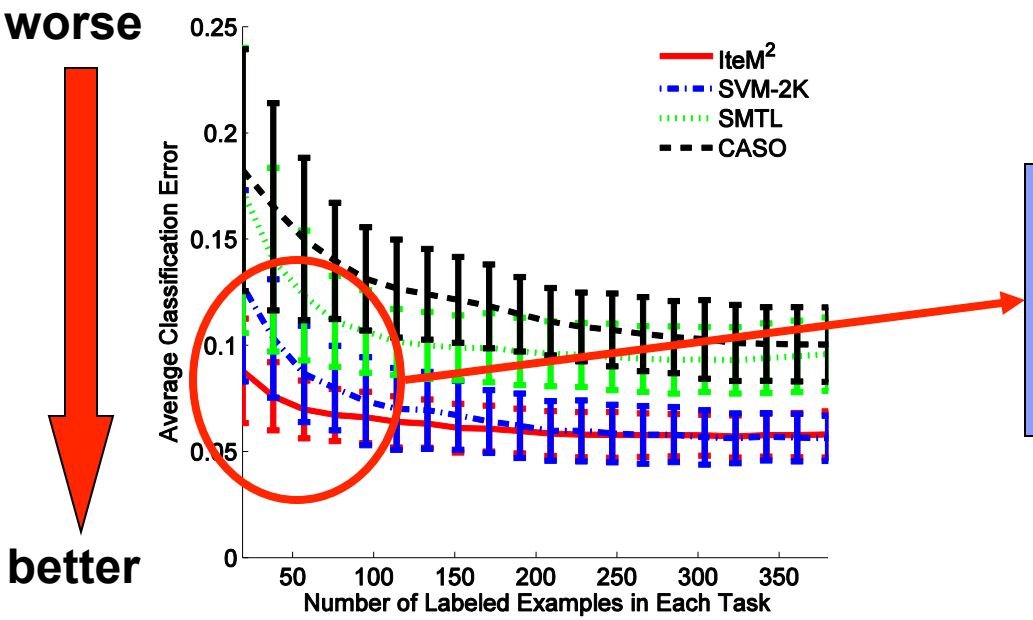


better

Consistently better than competitors

Multiple Tasks with Identical Views

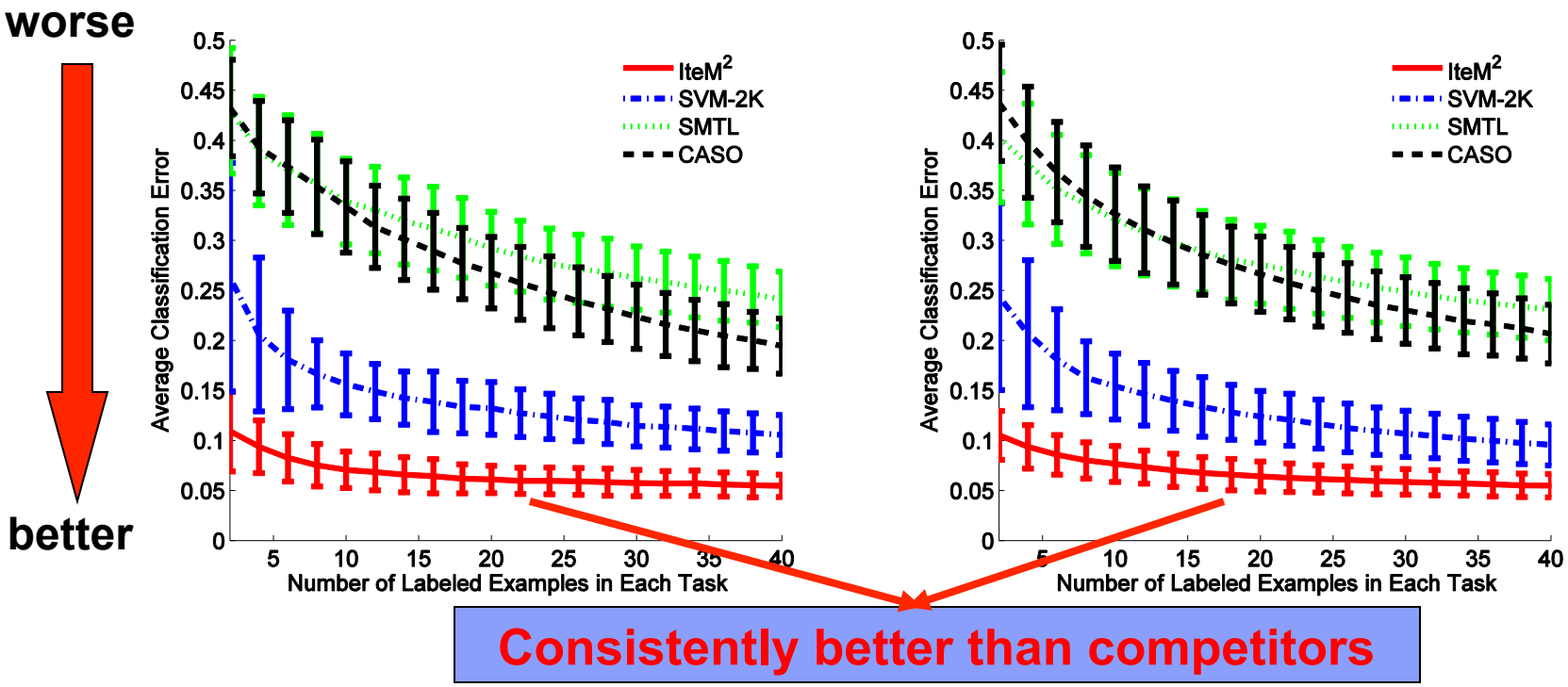
- Data set: WebKB data set
- Common views: words in the web page, anchor text, and title respectively



***IteM*² better than SVM-2K when labeled data is scarce**

Multiple Tasks and Non-identical Views

- Data set: ECML 2006 discovery challenge data set
- Common view: common vocabulary
- Task-specific view: task-specific vocabulary



Conclusion

- Multi-Task Multi-View (*M²TV*) Learning
  - Dual-heterogeneity: feature heterogeneity & task heterogeneity
- *GraM²*: a graph-based framework
  - Multi-partite graph capturing the relationship between examples and features
  - Objective function encodes similarity and consistency constraints
  - RKHS interpretation
- *IteM²*: an iterative procedure
  - Repeatedly update functions for examples and features