Multi-View Clustering with Constraint Propagation for Learning with an Incomplete Mapping Between Views

Eric Eaton
Bryn Mawr College*

Marie desJardins
University of Maryland Baltimore County

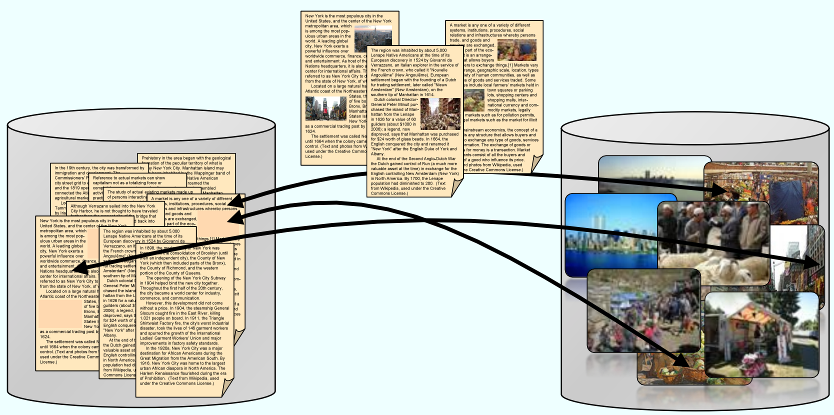
Sara Jacob
Lockheed Martin Advanced Technology Laboratories

This work was supported by internal funding from Lockheed Martin, NSF ITR #0325329, and a Graduate Fellowship from the Goddard Earth Sciences and Technology Center at UMBC.

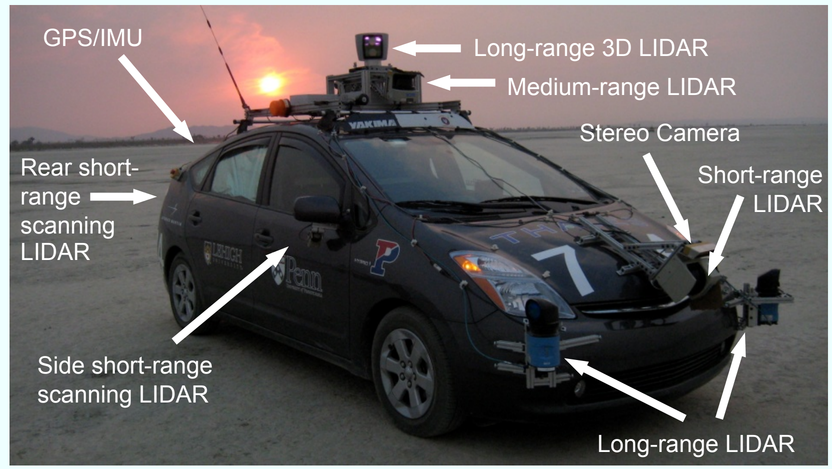
* The first author completed this work while at Lockheed Martin Advanced Technology Labs.

Introduction: Multi-view Learning

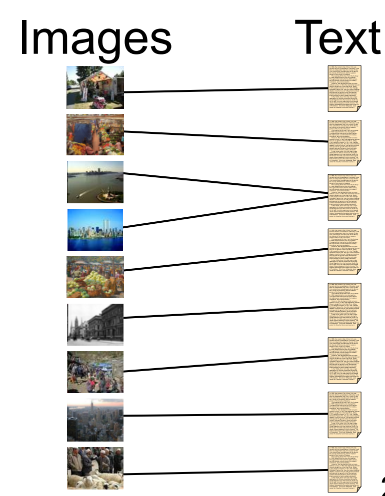
Multimodal Data Fusion and Retrieval
(field reports, websites)



Resolving Multiple Sensors

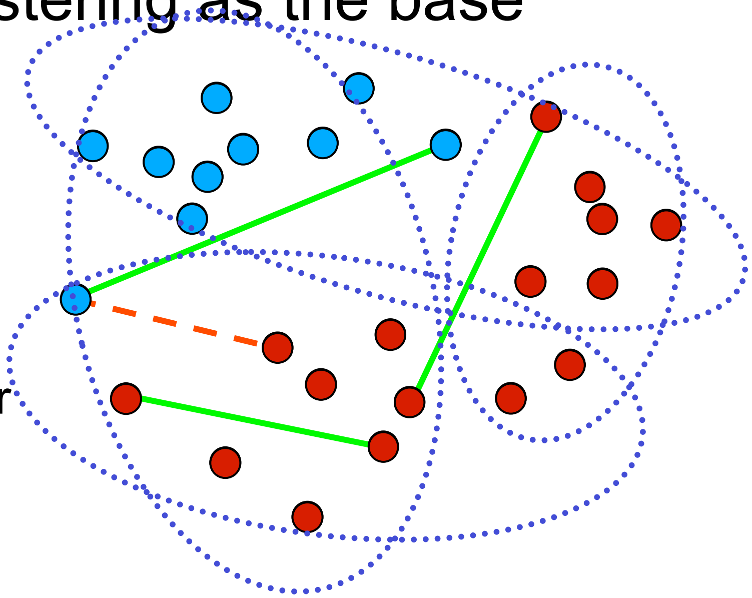


- Using multiple different views improves learning
- Most current methods assume a <u>complete</u> bipartite mapping between the views
  - This assumption is often unrealistic
  - Many applications yield only a partial mapping
- We focus on multi-view learning with a partial mapping between views



Background: Constrained Clustering

- Our approach uses constrained clustering as the base learning approach
  - Uses pairwise constraints to specify the relative cluster membership
    - Must-link constraint → same-cluster
    - Cannot-link constraint → different-cluster
  - Notation $\langle x_i, x_j, w, type \rangle \in \mathcal{C}$

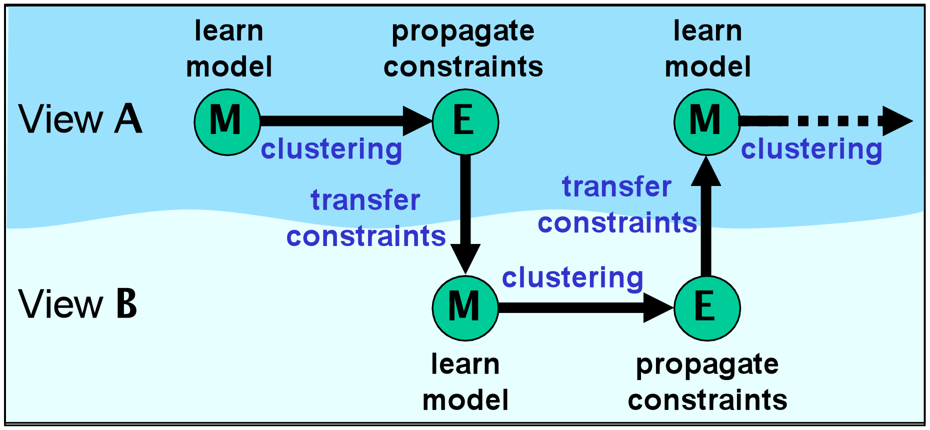


- PCK-Means Algorithm (Basu et al. 2002)
  - Incorporates constraints into K-Means objective function
  - Treats constraints as soft (can be violated with penalty **w**)
- MPCK-Means algorithm (Bilenko et al. 2004)
  - Also automatically learns distance metric for each cluster

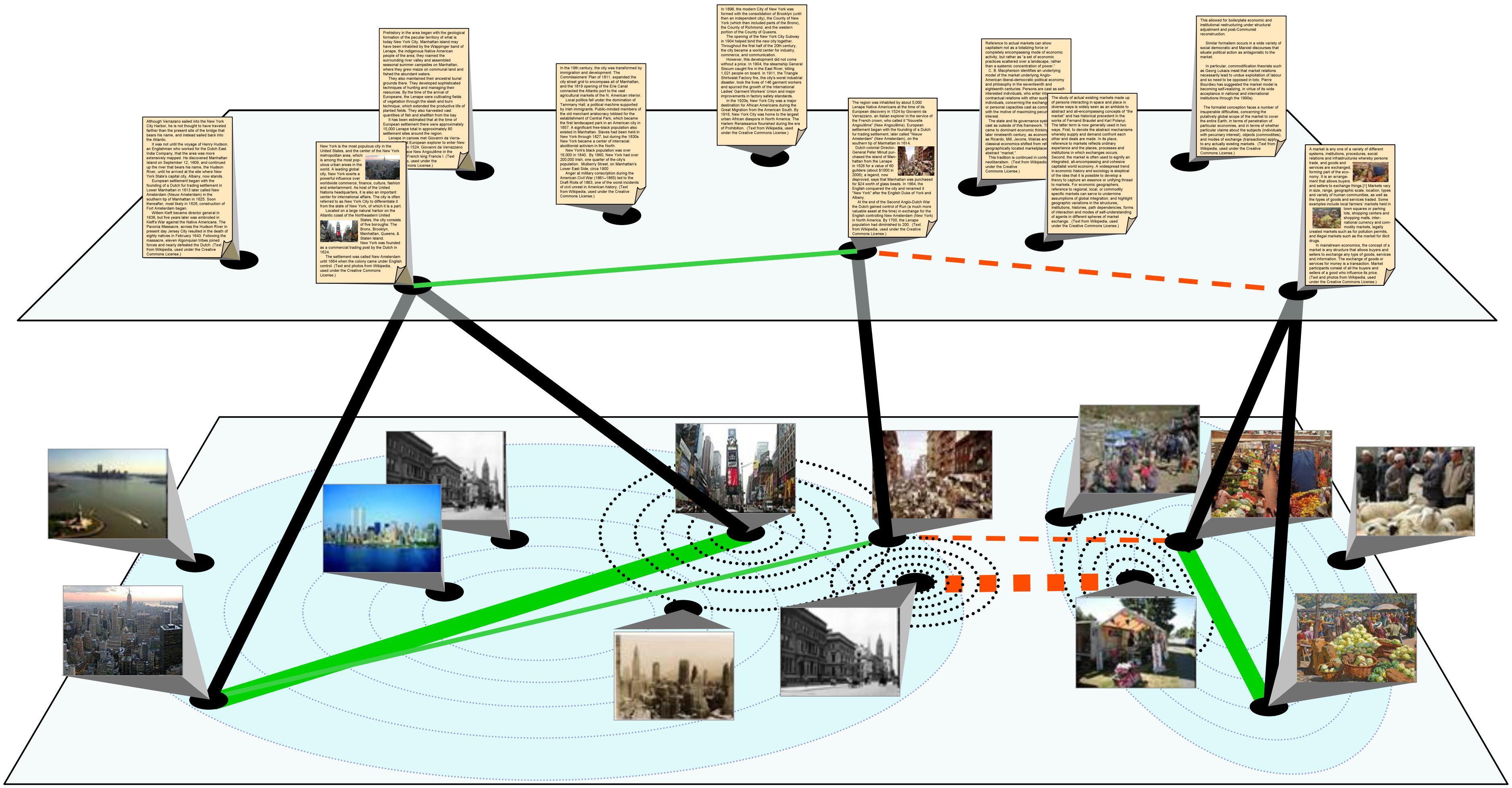
Our Approach

- Input: – Data $X^V = \{x_1^V, x_2^V, \ldots, x_{n_V}^V\} \subset \mathbb{R}^{d_V}$ for view V
  - – Bipartite mapping between views $\mathcal{R}^{A \times B} \subseteq X^A \times X^B$
  - – Set of constraints within each view $\mathcal{C}^A$ and $\mathcal{C}^B$

- Learn a cohesive clustering across views that respects the given constraints and (incomplete) mapping
  - – For each view:
    1.) Cluster the data, obtaining a model for the view
    2.) Propagate constraints within the view based on that model
    3.) Transfer those constraints across views to affect learning
  - – Repeat this process until convergence

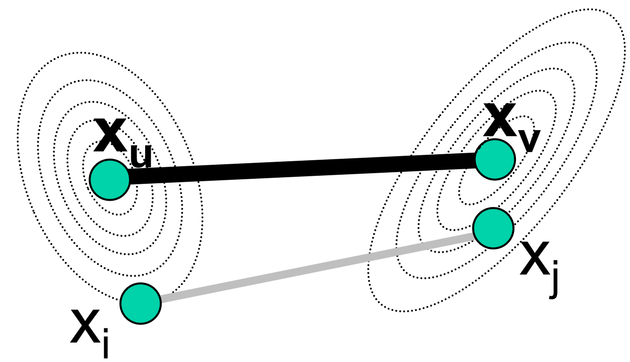


Multi-view Clustering with Constraint Propagation



Must-link   Cannot-link

Constraint Propagation



- Given constraint $\langle x_u, x_v, w \rangle \in \mathcal{C}$
- Infer constraint between x_i and x_j if they are sufficiently similar to $\langle x_u, x_v \rangle$ according to a local similarity measure
- Weight $\hat{w}$ of constraint $\langle x_i, x_j \rangle$ given by the radial basis function centered at $\begin{bmatrix} x_u \\ x_v \end{bmatrix}$ with covariance matrix shaped like clustering model:

$$\hat{w} = w \times \exp\left(-\frac{1}{2} \left\| \begin{bmatrix} x_i \\ x_j \end{bmatrix} - \begin{bmatrix} x_u \\ x_v \end{bmatrix} \right\|^2_{\Sigma_{uv}^{-1}} \right)$$

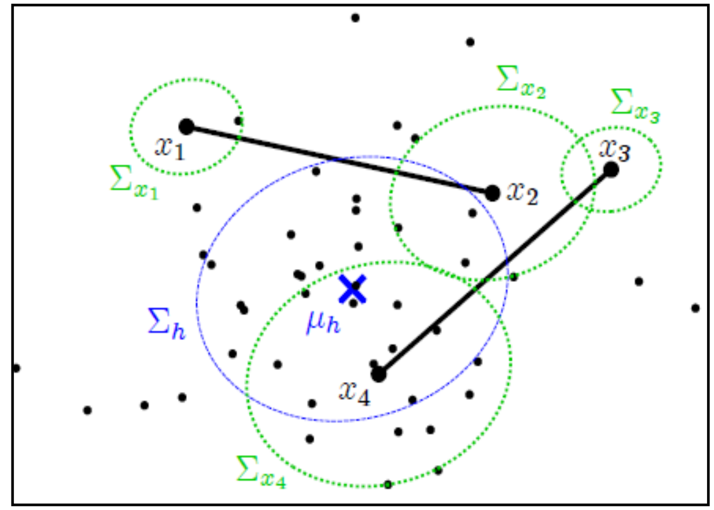
  – Each $x_i \in \mathbb{R}^{d_V}$, similarity measured in $\mathbb{R}^{2d_V}$

  – x_i assumed closest to x_u (same for x_j and x_v) since order matters

Constraint Propagation

From before: propagate constraint $\langle x_u, x_v, w \rangle \in \mathcal{C}$ to $\langle x_u, x_v \rangle$ with weight

$$\hat{w} = w \times \exp\left(-\frac{1}{2}\left\| \begin{bmatrix} x_i \\ x_j \end{bmatrix} - \begin{bmatrix} x_u \\ x_v \end{bmatrix} \right\|^2_{\Sigma_{uv}^{-1}}\right)$$



- Assuming independence between the endpoints yields

$$\hat{w} = w \times \exp\left(-\frac{1}{2}\|x_i - x_u\|^2_{\Sigma_u^{-1}}\right) \times \exp\left(-\frac{1}{2}\|x_j - x_v\|^2_{\Sigma_v^{-1}}\right)$$

  – The covariance matrix Σ_u controls the distance of propagation

  – Intuitively, constraints near the center of the cluster μ_h have high confidence and should be propagated a long distance

  – Idea: scale cluster covariance Σ_h by distance from centroid μ_h

$$\Sigma_u = \exp\left(-\frac{1}{2}\|x_u - \mu_h\|^2_{\Sigma_h^{-1}}\right) \times \Sigma_h$$

Multi-View Constraint Propagation Algorithm

Input: – Data $X^V = \{x_1^V, x_2^V, \ldots, x_{n_V}^V\} \subset \mathbb{R}^{d_V}$ for views A and B

– Bipartite mapping between views $\mathcal{R}^{A \times B} \subseteq X^A \times X^B$

– Set of constraints within each view $\mathcal{C}^A$ and $\mathcal{C}^B$

Initialize the propagated constraints $\mathcal{P}^A = \emptyset$, $\mathcal{P}^B = \emptyset$

Initialize constraint mapping functions $f_{A \to B}$, $f_{B \to A}$ from $\mathcal{R}^{A \times B}$

Repeat until convergence

for each view V (let U denote the opposing view)

1.) Form the unified set of constraints $\tilde{\mathcal{C}}^V = \mathcal{C}^V \overset{\max}{\bigcup} f_{U \to V}(\mathcal{P}^U)$

2.) M-step: Cluster view V using constraints $\tilde{\mathcal{C}}^V$

3.) E-step: Re-estimate the set of propagated constraints $\mathcal{P}^V$ using the updated clustering

end for

Extension to multiple views: $\tilde{\mathcal{C}}^V = \mathcal{C}^V \overset{\max}{\bigcup}{}_{U=1}^{D} f_{U \to V}(\mathcal{P}^U)$

Evaluation

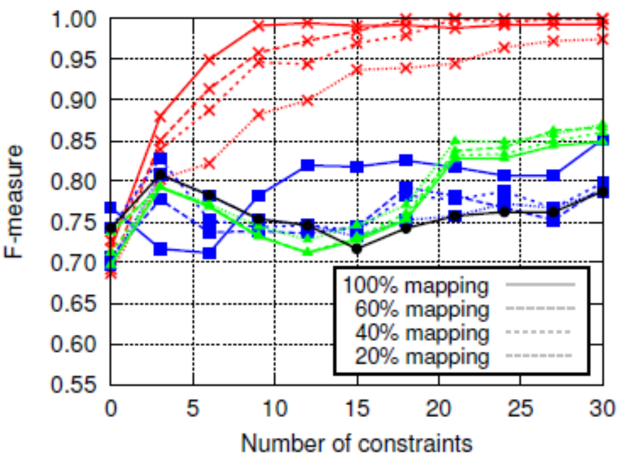
- Tested on a combination of synthetic and real data sets

Data Set Name	Description	Num Instances	Num Dimensions	Num Clusters	Propagation Threshold
Four Quadrants	Synthetic	200/200	2	2	0.75
Protein	Bioinformatics	67/49	20	3	0.5
Letters/Digits	Character Recognition	227/317	16	3	0.95
Rec/Talk (20 newsgroups)	Text Categorization	100/94	50	2	0.75

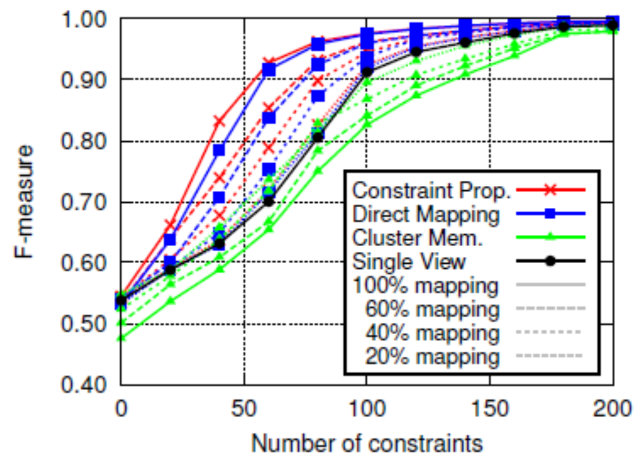
  – Constraint propagation works best in low-dimensions (due to curse of dimensionality), so we use the $\sqrt{d_V}$ spectral features

- Compare to:
  – *Direct Mapping*: equivalent to current methods for multi-view learning
  – *Cluster Membership*: infer constraints based on the current clustering
  – *Single View*: clustering each view in isolation

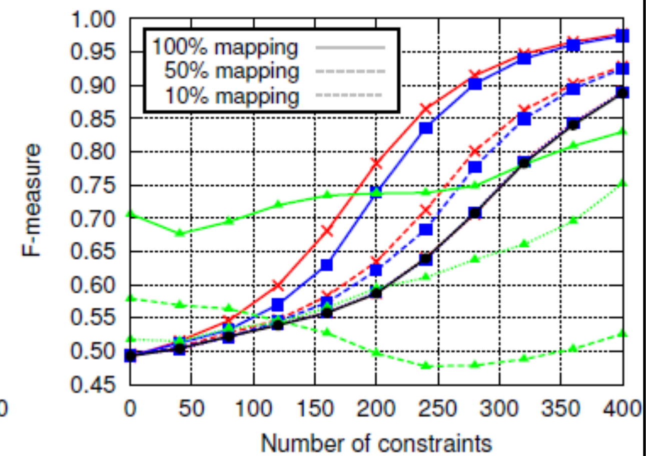
Results



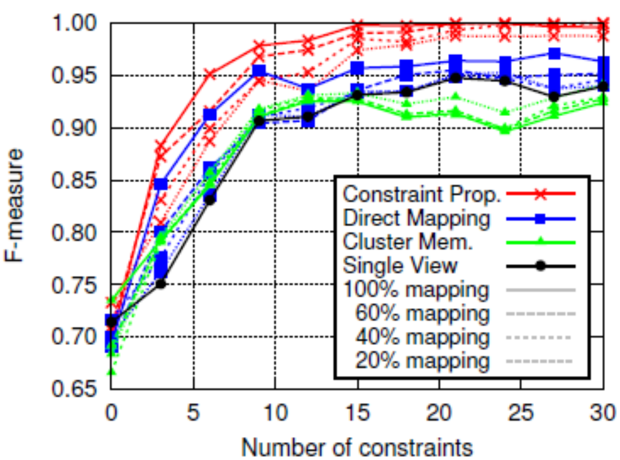
(a) Four Quadrants - PCK-Means



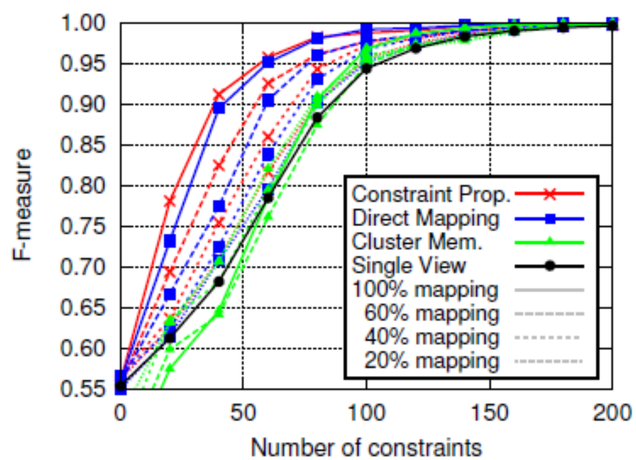
(b) Protein - PCK-Means



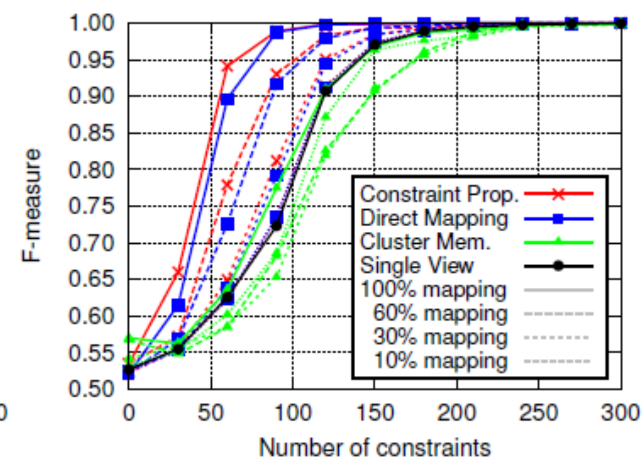
(c) Letters/Digits - PCK-Means



(d) Four Quadrants - MPCK-Means

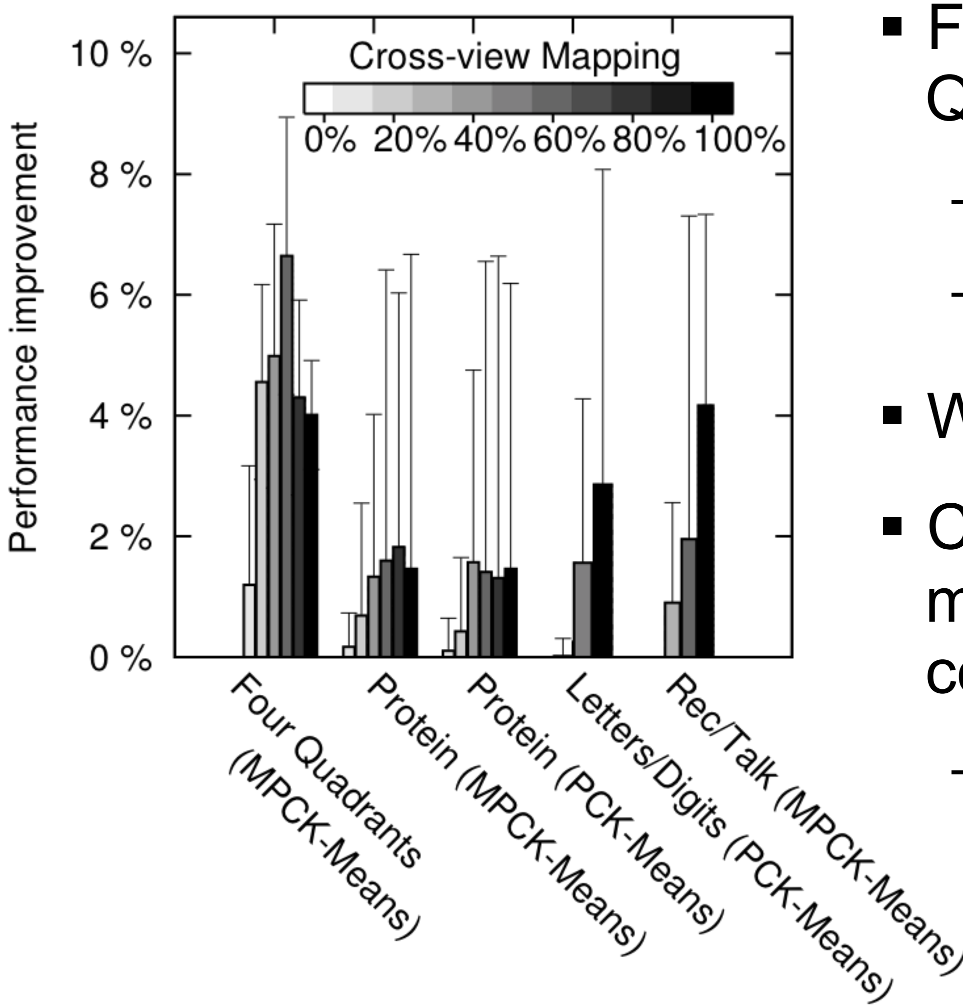


(e) Protein - MPCK-Means



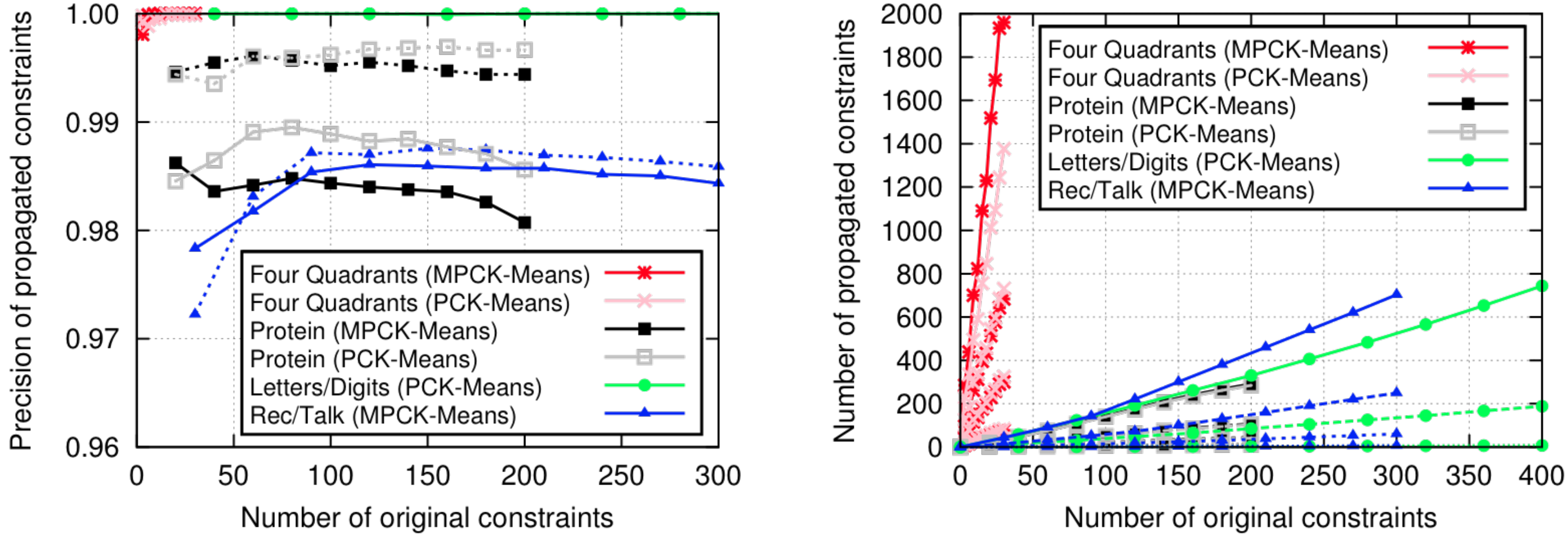
(f) Rec/Talk - MPCK-Means

Results: Improvement over Direct Mapping



- Figure omits results on Four Quadrants using PCK-Means
  - Average gains of 21.3%
  - Peak gains above 30%
- Whiskers show peak gains
- Constraint propagation still maintains a benefit even with a complete mapping
  - We hypothesize that it behaves similarly to spatial constraints (Klein et al., 2002) by warping the underlying space to improve performance

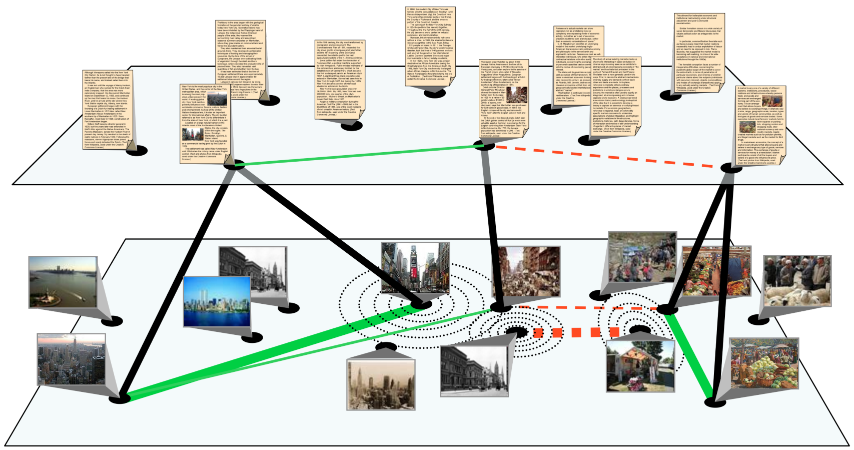
Results: Effects of Constraint Propagation



- Few incorrect constraints are inferred by the propagation
- Constraint propagation works slightly better for cannot-link constraints than must-link constraints
  - *Counting Argument*: there are many more chances for a cannot-link constraint to be correctly propagated than a must-link constraint

Conclusion and Future Work

- Constraint propagation improves multi-view constrained clustering under a partial mapping between views
- Provides the ability for the user to interact with one view, and for the interaction to affect the other views
  - E.g., the user constrains images, and it affects the clustering of texts
- Future work:
  - Inferring mappings from alignment of manifolds underlying views
  - Scaling up multi-view learning to many views, each with very few connections to other views
  - Using transfer to improve learning across distributions under a partial mapping between views



Thank You!
Questions?

Eric Eaton
eeaton@cs.brynmawr.edu

This work was supported by internal funding from Lockheed Martin, NSF ITR #0325329, and a Graduate Fellowship from the Goddard Earth Sciences and Technology Center at UMBC.

References

Asuncion, A. and D. Newman. UCI machine learning repository. Available at http://www.ics.uci.edu/mlearn/MLRepository.html.

Bar-Hillel, A.; T. Hertz; N. Shental; and D. Weinshall. 2005. Learning a Mahalanobis metric from equivalence constraints. Journal of Machine Learning Research, 6:937-965.

Basu, S. 2005. Semi-Supervised Clustering: Probabilistic Models, Algorithms, and Experiments. PhD thesis, University of Texas at Austin.

Basu, S.; A. Banerjee; and R. Mooney. 2002. Semi-supervised clustering by seeding. In Proceedings of ICML-02, pages 19-26. Morgan Kaufman.

Basu, S.; A. Banerjee; and R. J. Mooney. 2004. Active semi-supervision for pairwise constrained clustering. In Proceedings of ICDM-04, pages 333{344, 2004. SIAM.

Bickel, S. and T. Scheer. 2004. Multi-view clustering. In Proceedings of IEEE ICDM-04, pages 19-26, Washington, DC. IEEE Computer Society.

Bilenko, M.; S. Basu; and R. J. Mooney. 2004. Integrating constraints and metric learning in semi-supervised clustering. In Proceedings of ICML-04, pages 81-88. ACM.

Blum, A. and T. Mitchell. 1998. Combining labeled and unlabeled data with co-training. In Proceedings of COLT-98, pages 92-100. Morgan Kaufmann.

Chaudhuri, K.; S. M. Kakade; K. Livescu; and K. Sridharan. 2009. Multi-view clustering via canonical correlation analysis. In Proceedings of ICML-09, pages 129-136, New York. ACM.

Chung, F. R. K. 1994. Spectral Graph Theory. Number 92 in CBMS Regional Conference Series in Mathematics. American Mathematical Society, Providence, RI.

Dean, J. and S. Ghemawat. 2008. MapReduce: simplied data processing on large clusters. Communications of the ACM, 51(1):107-113.

Klein, D.; S. D. Kamvar; and C. D. Manning. From instance-level constraints to space-level constraints. In Proceedings of ICML-02, pages 307-314. Morgan Kaufman.

Ng, A. Y.; M. I. Jordan; and Y. Weiss. 2001. On spectral clustering: Analysis and an algorithm. In NIPS 14, pages 849-856. MIT Press.

Nigam, K. and R. Ghani. 2000. Analyzing the effectiveness and applicability of co-training. In Proceedings of CIKM-00, pages 86-93, New York, NY. ACM.

Rennie, J. 2003. 20 Newsgroups data set, sorted by date. Available online at http://www.ai.mit.edu/~jrennie/20Newsgroups/.

Wagstaff, K.; C. Cardie; S. Rogers; and S. Schroedl. 2001. Constrained k-means clustering with background knowledge. In Proceedings of ICML-01, pages 577-584. Morgan Kaufmann.

Wagstaff, K. 2002. Intelligent Clustering with Instance-Level Constraints. PhD thesis, Cornell University.

Witten, I.H. and E. Frank. 2005. Data Mining: Practical Machine Learning Tools and Techniques, 2nd edition. Morgan Kaufmann.

Xing, E. P. ; A. Y. Ng; M. I. Jordan; and S. Russell. 2003. Distance metric learning, with application to clustering with side-information. Advances in Neural Information Processing Systems, 15:505-512.