Partial Multi-View Clustering Using Graph Regularized NMF

Nishant Rai – IIT Kanpur
Sumit Negi – Amazon Development Center
Santanu Chaudhury – IIT Delhi
Om Deshmukh – Xerox Research Centre India

ICPR '16, Cancun, Mexico

INTRODUCTION

Task

- Multi-view learning algorithms aim at exploiting the complementary information present in different views.
- Most methods assume that each example appears in all views.
- Such an assumption is too restrictive in real world settings.
- **Task:** Classification and clustering when (possibly) every view suffers from missing information.

Related Work

- **Multi-View K-Means:** Extension of K means to multiple views after modifying cost function
- **Multi-View Spectral Clustering:** Such approaches exploit some similarity measure between objects [5] [6]
- **Multi-View Subspace Clustering:** Such methods assume that the views are generated from a common subspace [3]
- **Multi-view clustering with partial examples:** Addresses the scenario where every view might suffer from missing information [4]

MOTIVATION

Non Negative Matrix Factorization

In NMF we choose factors U and V of X such that the following objective is minimized,

$$\min_{U,V} \left\|X - UV^T\right\|_F^2 \qquad \text{s.t. } U \geq 0,\ V \geq 0$$

Advantages of Non Negative Matrix Factorization,

- Due to the non negativity constraints, NMF produces an 'additive' parts-based representation of the data matrix X.
- Consequently, the factors of X are generally naturally sparse.
- Leads to impressive benefits in terms of interpretability of its factors.

Intuition

- Samples in different views should be assigned to similar clusters [2]
  - $\mu_i \left\| V_i Q_i - V^* \right\|_F^2$ term added to loss
  - Q_i [2] added for normalization[1]
- Respect the intrinsic geometrical representation of each view.
  - $\lambda_i Tr(V_i^T L_i V_i)$ term[1] added for Graph Regularization [1].
- In order to support partial view setup[2], we introduce view-to-instance mappings P_i[1] for each view.

[1] Details in the paper

[2] Not all instances represented in all views

Loss function

- Based on the previous discussions, the loss function is,

$$\min_{U_i, V_i, V^*} \sum_{i=1}^{v} \left(\left\| X_i - U_i V_i^T \right\|_F^2 + \mu_i \left\| V_i Q_i - V_{P_i}^* \right\|_F^2 + \lambda_i Tr(V_i^T L_i V_i) \right)$$

$$s.t. \;\; U_i \geq 0, \; V_i \geq 0, \; \forall i \;\; s.t. \; 1 \leq i \leq v$$

METHODOLOGY

Our Approach

- As discussed previously,

$$\min_{U_i,V_i,V^*} \sum_{i=1}^{v} \left(\left\| X_i - U_i V_i^T \right\|_F^2 + \mu_i \left\| V_i Q_i - V_{P_i}^* \right\|_F^2 + \lambda_i Tr(V_i^T L_i V_i) \right)$$

$$s.t. \;\; U_i \geq 0,\; V_i \geq 0,\; \forall i \;\; s.t.\; 1 \leq i \leq v$$

- This is non-convex optimization problem and is thus difficult to optimize.
- Since the loss is convex in each variable individually, we separately optimize the loss with w.r.t each variable.
- Such an alternate optimization is repeated till convergence.

Our Algorithm[1]

Algorithm 1: Graph Regularized Partial Multi-View Clustering Algorithm (GPMVC)

Input : Nonnegative data matrix $X_1, \ldots, X_v$; Parameters $\lambda_1, \mu_1, \ldots, \lambda_v, \mu_v$; Number of clusters K; *View-to-Instance* mapping P; *Inverse-Mapping* P^{-}

Output: Basis Matrices $U_1, \ldots, U_v$; Coefficient Matrices $V_1, \ldots, V_v$ and Consensus Matrix V^*

Construct Graph Laplacians L_i for each view;
Normalize X_i such that $\|X_i\|_1 = 1$ for each view ;
Initialize U_i, V_i and V^* by Eq. 5, Eq. 9;
repeat
    for $i \leftarrow 1$ **to** v **do**
        repeat
            Fix V^* and V_i, update U_i by Eq. 7;
            Fix V^* and U_i, update V_i by Eq. 8;
            Normalize U_i, V_i using Q_i;
        until *objective function in Eq. 4 converges*;
    end
    Fix U, V update V^* by Eq. 9;
until *objective function in Eq. 4 converges*;

---

[1] Details regarding the equations and algorithm present in paper

RESULTS

Dataset Description

We use five publicly available text and image datasets for our experiments. The dataset statistics are described below,

Datasets	Size	# Views	# Clusters
3Sources	169	3	6
Digit	2000	5	10
ORL	400	2	40
BBCSports	282	3	5
Cora	2708	2	7

Table: Dataset statistics

Partial View Dataset Construction

- Randomly select a fraction of the instances to be *partial examples* i.e., they are present in only one of the views, remaining instances are complete.
- The incomplete instances (*partial examples*) are equally shared amongst all the views. The Partial Example Ratio (PER) dictates the fraction of partial examples.
- We later change this assumption (Of equally sharing partial examples) by introducing a skew factor (SF).

[1] To reduce bias, we report the average results on 10 versions of the dataset for each PER.

Quantitative Results

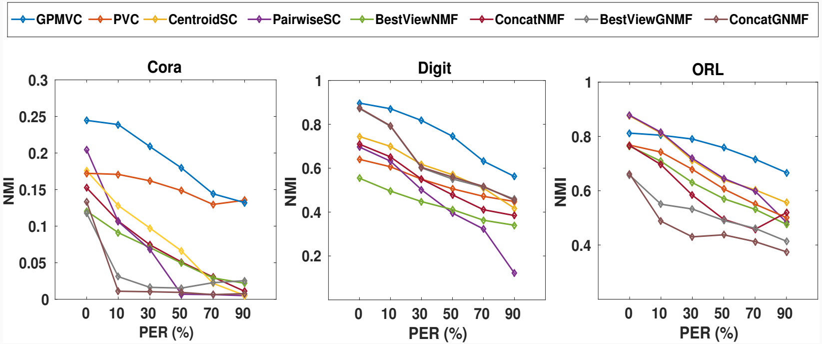

Figure: NMI values vs. PER for Two View datasets

Quantitative Results

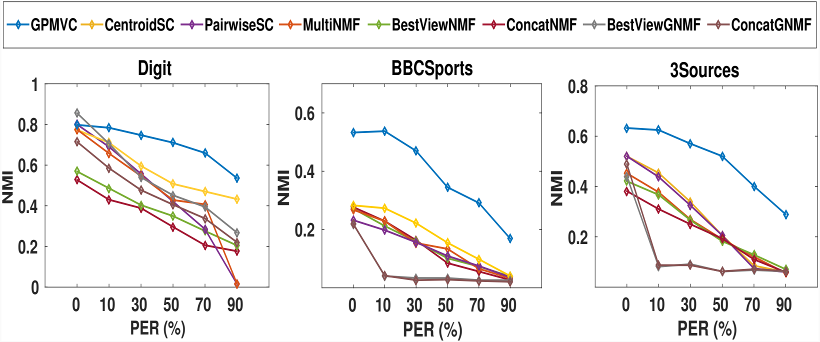

Figure: NMI values vs. PER for Multi-View datasets

Qualitative Results

Effect of Skew: The Skew Factor (SF) controls how the partial examples are distributed between the two views.

SF(%)	*10*		*30*		*70*		*90*	
PER (%)	GPMVC	PVC	GPMVC	PVC	GPMVC	PVC	GPMVC	PVC
10	**0.900**	0.632	**0.886**	0.632	**0.875**	0.608	**0.882**	0.604
30	**0.880**	0.630	**0.866**	0.629	**0.825**	0.527	**0.808**	0.507
50	**0.828**	0.614	**0.789**	0.614	**0.728**	0.482	**0.733**	0.453
70	**0.811**	0.582	**0.688**	0.582	**0.655**	0.446	**0.679**	0.445
90	**0.748**	0.555	**0.637**	0.555	**0.588**	0.460	**0.638**	0.493

Table: NMI on Digit (2 view)

[1] PVC: Multi View Clustering with Partial Examples [4]
[2] GPMVC: Our Proposed Approach

REFERENCES

References I

Deng Cai et al. "Graph Regularized Nonnegative Matrix Factorization for Data Representation". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 33.8 (2011), pp. 1548–1560. DOI: 10.1109/TPAMI.2010.231.

Jing Gao et al. "Multi-View Clustering via Joint Nonnegative Matrix Factorization." In: *SDM*. SIAM, 2013, pp. 252–260. ISBN: 978-1-61197-283-2.

Derek Greene and Padraig Cunningham. "A Matrix Factorization Approach for Integrating Multiple Data Views." In: *ECML/PKDD (1)*. Ed. by Wray L. Buntine et al. Vol. 5781. Lecture Notes in Computer Science. Springer, Aug. 31, 2009, pp. 423–438. ISBN: 978-3-642-04179-2.

References II

- Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. "Partial Multi-View Clustering." In: *AAAI*. Ed. by Carla E. Brodley and Peter Stone. AAAI Press, 2014, pp. 1968–1974. ISBN: 978-1-57735-661-5.
- V. R. de Sa. "Spectral clustering with two views". In: *ICML Workshop on Learning With Multiple Views*. 2005.
- Dengyong Zhou and Christopher J. C. Burges. "Spectral clustering and transductive learning with multiple views." In: *ICML*. Ed. by Zoubin Ghahramani. Vol. 227. ACM International Conference Proceeding Series. ACM, 2007, pp. 1159–1166. ISBN: 978-1-59593-793-3.