Co-regularized Spectral Clustering with Multiple Kernels

Abhishek Kumar[‡], Piyush Rai[*], Hal Daumé III[‡]

[*] School of Computing, University of Utah
[‡] Dept. of Computer Science, University of Maryland

December 11, 2010

Learning with Multiple Views

- Many datasets admit multiple representations. For example:
  - Webpages: Page-text, hyperlinks, social tags, etc.
  - Images: Different forms of extracted features (Pixels, Fourier coefficients, etc)
- Each representation is a **view** of the data
- Each view **individually** is good enough to learn from, given *enough* data

Learning with Multiple Views

- Many datasets admit multiple representations. For example:
  - Webpages: Page-text, hyperlinks, social tags, etc.
  - Images: Different forms of extracted features (Pixels, Fourier coefficients, etc)
- Each representation is a **view** of the data
- Each view **individually** is good enough to learn from, given *enough* data
- Multiview Learning: Exploit multiple views to do even better
  - Require lesser data to learn
  - .. and hopefully learn better

Learning with Multiple Views

- Many datasets admit multiple representations. For example:
  - Webpages: Page-text, hyperlinks, social tags, etc.
  - Images: Different forms of extracted features (Pixels, Fourier coefficients, etc)
- Each representation is a **view** of the data
- Each view **individually** is good enough to learn from, given *enough* data
- Multiview Learning: Exploit multiple views to do even better
  - Require lesser data to learn
  - .. and hopefully learn better
- vis-a-vis Multiple Kernel Learning (MKL)
  - Each view can be used to define a similarity graph or a kernel
  - In a kernel based setting, MKL and Multiview Learning are synonymous

Co-regularization based Multiview Learning

- Idea: Enforcing agreement between learners defined over different views
- Typically used in semi-supervised learning (e.g., Co-training)
  - Two hypotheses f_1 and f_2 learned on views $\mathcal{V}_1$ and $\mathcal{V}_2$
  - Enforce agreement on unlabeled data ($f_1(x) = f_2(x)$)
  - Requires lesser labeled data to learn (and often learns better)

Co-regularization based Multiview Learning

- Idea: Enforcing agreement between learners defined over different views
- Typically used in semi-supervised learning (e.g., Co-training)
  - Two hypotheses f_1 and f_2 learned on views $\mathcal{V}_1$ and $\mathcal{V}_2$
  - Enforce agreement on unlabeled data ($f_1(x) = f_2(x)$)
  - Requires lesser labeled data to learn (and often learns better)
- This talk: Using co-regularization for clustering (an unsupervised problem)
  - In the context of Spectral Clustering (a kernel based clustering algorithm)

Co-regularization based Multiview Learning

- Idea: Enforcing agreement between learners defined over different views
- Typically used in semi-supervised learning (e.g., Co-training)
  - Two hypotheses f_1 and f_2 learned on views $\mathcal{V}_1$ and $\mathcal{V}_2$
  - Enforce agreement on unlabeled data ($f_1(x) = f_2(x)$)
  - Requires lesser labeled data to learn (and often learns better)
- This talk: Using co-regularization for clustering (an unsupervised problem)
  - In the context of Spectral Clustering (a kernel based clustering algorithm)
  - Idea: Enforce clusterings from multiple views to agree with each other
  - Note: each view corresponds to a kernel

Co-regularization based Multiview Learning

- Idea: Enforcing agreement between learners defined over different views
- Typically used in semi-supervised learning (e.g., Co-training)
  - Two hypotheses f_1 and f_2 learned on views $\mathcal{V}_1$ and $\mathcal{V}_2$
  - Enforce agreement on unlabeled data $(f_1(x) = f_2(x))$
  - Requires lesser labeled data to learn (and often learns better)
- This talk: Using co-regularization for clustering (an unsupervised problem)
  - In the context of Spectral Clustering (a kernel based clustering algorithm)
  - Idea: Enforce clusterings from multiple views to agree with each other
  - Note: each view corresponds to a kernel
- (As we will see) In our case, this is akin to combining kernels

Spectral Clustering

- Based on spectral decomposition of the Graph Laplacian of the data
- Theoretically well motivated, can learn arbitrary shaped clusters
- Some notations:
  - $\mathbf{K}$: $N \times N$ kernel matrix of data $\mathbf{X} \in \mathbb{R}^{N \times D}$
  - $\mathbf{K}_{ij}$ similarity between examples i and j
  - $\mathbf{D}$: diagonal matrix with $\mathbf{D}_{ii} = \sum_j \mathbf{K}_{ij}$
  - The normalized graph Laplacian $\mathcal{L} = \mathbf{D}^{-1/2}\mathbf{K}\mathbf{D}^{-1/2}$
- Given the graph Laplacian $\mathcal{L} \in \mathbb{R}^{N \times N}$, we seek K partitions of the data
  - via spectral decomposition of $\mathcal{L}$

Spectral Clustering (Contd.)

- The spectral clustering objective (Ng *et al*, NIPS 2002):

$$\max_{\mathbf{U}\in\mathbb{R}^{N\times K}} tr(\mathbf{U}^T\mathcal{L}\mathbf{U}) \qquad \text{s.t.} \quad \mathbf{U}^T\mathbf{U} = I$$

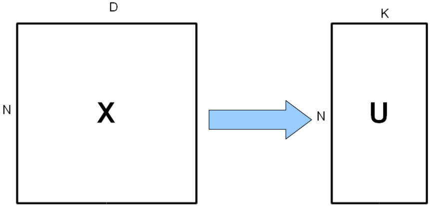
- An eigenvalue problem: Amounts to finding the K top eigenvectors of $\mathcal{L}$

Spectral Clustering (Contd.)

- The spectral clustering objective (Ng *et al*, NIPS 2002):

$$\max_{\mathbf{U} \in \mathbb{R}^{N \times K}} tr(\mathbf{U}^T \mathcal{L} \mathbf{U}) \qquad \text{s.t.} \quad \mathbf{U}^T\mathbf{U} = I$$

- An eigenvalue problem: Amounts to finding the K top eigenvectors of $\mathcal{L}$
- Can think of **U** as a new representation of **X**
  - $\mathbf{U}_i$ corresponds to $\mathbf{X}_i$ (the i^{th} example)

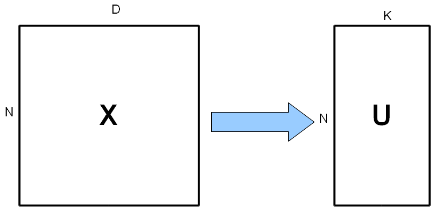


Spectral Clustering (Contd.)

- The spectral clustering objective (Ng *et al*, NIPS 2002):

$$\max_{\mathbf{U}\in\mathbb{R}^{N\times K}} tr(\mathbf{U}^T\mathcal{L}\mathbf{U}) \qquad \text{s.t.} \quad \mathbf{U}^T\mathbf{U} = I$$

- An eigenvalue problem: Amounts to finding the K top eigenvectors of $\mathcal{L}$
- Can think of $\mathbf{U}$ as a new representation of $\mathbf{X}$
    - $\mathbf{U}_i$ corresponds to $\mathbf{X}_i$ (the i^{th} example)



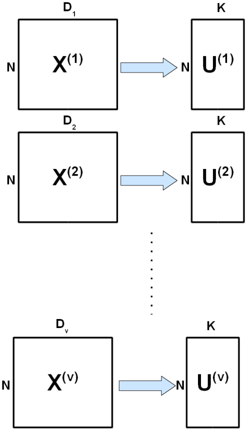
- **Final step:** normalize rows of $\mathbf{U}$ and run K-means

Spectral Clustering with Multiple Views

- We have access to multiple views of the data
- Let $\mathbf{X}^{(v)} = \{\mathbf{x}_1^{(v)}, \mathbf{x}_2^{(v)}, \ldots, \mathbf{x}_N^{(v)}\}$ denote the data in view v
- Denote the corresponding graph Laplacian by $\mathcal{L}^{(v)}$
- The spectral clustering objective for each individual view v:

$$\max_{\mathbf{U}^{(v)} \in \mathbb{R}^{N \times K}} tr(\mathbf{U}^{(v)^T} \mathcal{L}^{(v)} \mathbf{U}^{(v)}), \text{s.t.} \mathbf{U}^{(v)^T} \mathbf{U}^{(v)} = I$$

- Co-regularization: Enforce the **U**'s from all the views to look similar to each other *in some sense* (to be defined later)



Co-regularized Spectral Clustering

- Co-regularized spectral clustering objective for 2 views v and w

$$\max_{\mathbf{U}^{(v)}\in\mathbb{R}^{N\times K},\mathbf{U}^{(w)}\in\mathbb{R}^{N\times K}} \underbrace{tr(\mathbf{U}^{(v)^T}\mathcal{L}^{(v)}\mathbf{U}^{(v)})}_{view\ v\ objective} + \underbrace{tr(\mathbf{U}^{(w)^T}\mathcal{L}^{(w)}\mathbf{U}^{(w)})}_{view\ w\ objective} \underbrace{+\lambda D(\mathbf{U}^{(v)},\mathbf{U}^{(w)})}_{co-regularization\ term}$$

$$\text{s.t.}\quad \mathbf{U}^{(v)^T}\mathbf{U}^{(v)} = I,\ \mathbf{U}^{(w)^T}\mathbf{U}^{(w)} = I$$

- Note: Extension to more than 2 views in a likewise manner
- What should the co-regularizer $D(.,.)$ look like?

Co-regularized Spectral Clustering

- Co-regularized spectral clustering objective for 2 views v and w

$$\max_{\mathbf{U}^{(v)} \in \mathbb{R}^{N \times K}, \mathbf{U}^{(w)} \in \mathbb{R}^{N \times K}} \underbrace{tr(\mathbf{U}^{(v)^T} \mathcal{L}^{(v)} \mathbf{U}^{(v)})}_{view\ v\ objective} + \underbrace{tr(\mathbf{U}^{(w)^T} \mathcal{L}^{(w)} \mathbf{U}^{(w)})}_{view\ w\ objective} \underbrace{+\lambda D(\mathbf{U}^{(v)}, \mathbf{U}^{(w)})}_{co-regularization\ term}$$

$$\text{s.t.} \quad \mathbf{U}^{(v)^T} \mathbf{U}^{(v)} = I, \ \mathbf{U}^{(w)^T} \mathbf{U}^{(w)} = I$$

- Note: Extension to more than 2 views in a likewise manner
- What should the co-regularizer $D(.,.)$ look like?
- Intuition: Each view should lead to the same clustering in **U** space

Co-regularized Spectral Clustering

- Co-regularized spectral clustering objective for 2 views v and w

$$\max_{\mathbf{U}^{(v)}\in\mathbb{R}^{N\times K},\mathbf{U}^{(w)}\in\mathbb{R}^{N\times K}} \underbrace{tr(\mathbf{U}^{(v)^T}\mathcal{L}^{(v)}\mathbf{U}^{(v)})}_{view\ v\ objective} + \underbrace{tr(\mathbf{U}^{(w)^T}\mathcal{L}^{(w)}\mathbf{U}^{(w)})}_{view\ w\ objective} \underbrace{+\lambda D(\mathbf{U}^{(v)},\mathbf{U}^{(w)})}_{co-regularization\ term}$$

$$\text{s.t.}\quad \mathbf{U}^{(v)^T}\mathbf{U}^{(v)} = I,\ \mathbf{U}^{(w)^T}\mathbf{U}^{(w)} = I$$

- Note: Extension to more than 2 views in a likewise manner
- What should the co-regularizer $D(.,.)$ look like?
- Intuition: Each view should lead to the same clustering in $\mathbf{U}$ space
- Condition: Kernels defined over $\mathbf{U}$ should look similar for all views
  - Implies high degree of alignment between $\mathbf{K}_{\mathbf{U}^{(v)}}$ and $\mathbf{K}_{\mathbf{U}^{(w)}}$

Co-regularized Spectral Clustering

- Co-regularized spectral clustering objective for 2 views v and w

$$\max_{\mathbf{U}^{(v)} \in \mathbb{R}^{N \times K}, \mathbf{U}^{(w)} \in \mathbb{R}^{N \times K}} \underbrace{tr(\mathbf{U}^{(v)^T} \mathcal{L}^{(v)} \mathbf{U}^{(v)})}_{view\ v\ objective} + \underbrace{tr(\mathbf{U}^{(w)^T} \mathcal{L}^{(w)} \mathbf{U}^{(w)})}_{view\ w\ objective} \underbrace{+\lambda D(\mathbf{U}^{(v)}, \mathbf{U}^{(w)})}_{co-regularization\ term}$$

$$\text{s.t.} \quad \mathbf{U}^{(v)^T} \mathbf{U}^{(v)} = I, \ \mathbf{U}^{(w)^T} \mathbf{U}^{(w)} = I$$

- Note: Extension to more than 2 views in a likewise manner
- What should the co-regularizer $D(.,.)$ look like?
- Intuition: Each view should lead to the same clustering in **U** space
- Condition: Kernels defined over **U** should look similar for all views
  - Implies high degree of alignment between $\mathbf{K}_{\mathbf{U}^{(v)}}$ and $\mathbf{K}_{\mathbf{U}^{(w)}}$
- We use a linear kernel in **U** space: $\mathbf{K}_{\mathbf{U}^{(v)}} = \mathbf{U}^{(v)} \mathbf{U}^{(v)^T}$

Co-regularized Spectral Clustering

- Co-regularized spectral clustering objective for 2 views v and w

$$\max_{\mathbf{U}^{(v)} \in \mathbb{R}^{N \times K}, \mathbf{U}^{(w)} \in \mathbb{R}^{N \times K}} \underbrace{tr(\mathbf{U}^{(v)^T} \mathcal{L}^{(v)} \mathbf{U}^{(v)})}_{view\ v\ objective} + \underbrace{tr(\mathbf{U}^{(w)^T} \mathcal{L}^{(w)} \mathbf{U}^{(w)})}_{view\ w\ objective} \underbrace{+\lambda D(\mathbf{U}^{(v)}, \mathbf{U}^{(w)})}_{co-regularization\ term}$$

$$\text{s.t.} \quad \mathbf{U}^{(v)^T} \mathbf{U}^{(v)} = I, \ \mathbf{U}^{(w)^T} \mathbf{U}^{(w)} = I$$

- Note: Extension to more than 2 views in a likewise manner
- What should the co-regularizer $D(.,.)$ look like?
- Intuition: Each view should lead to the same clustering in $\mathbf{U}$ space
- Condition: Kernels defined over $\mathbf{U}$ should look similar for all views
  - Implies high degree of alignment between $\mathbf{K}_{\mathbf{U}^{(v)}}$ and $\mathbf{K}_{\mathbf{U}^{(w)}}$
- We use a linear kernel in $\mathbf{U}$ space: $\mathbf{K}_{\mathbf{U}^{(v)}} = \mathbf{U}^{(v)} \mathbf{U}^{(v)^T}$
- Alignment measured by the trace of the product of $\mathbf{K}_{\mathbf{U}^{(v)}}$ and $\mathbf{K}_{\mathbf{U}^{(w)}}$

$$D(\mathbf{U}^{(v)}, \mathbf{U}^{(w)}) = tr(\mathbf{K}_{\mathbf{U}^{(v)}} \mathbf{K}_{\mathbf{U}^{(w)}}) = tr(\mathbf{U}^{(v)} \mathbf{U}^{(v)^T} \mathbf{U}^{(w)} \mathbf{U}^{(w)^T})$$

Co-regularized Spectral Clustering

- The objective becomes:

$$\max_{\mathbf{U}^{(v)} \in \mathbb{R}^{N \times K}, \mathbf{U}^{(w)} \in \mathbb{R}^{N \times K}} \underbrace{tr(\mathbf{U}^{(v)^T} \mathcal{L}^{(v)} \mathbf{U}^{(v)})}_{view\ \ v\ \ objective} + \underbrace{tr(\mathbf{U}^{(w)^T} \mathcal{L}^{(w)} \mathbf{U}^{(w)})}_{view\ \ w\ \ objective} \underbrace{+\lambda tr(\mathbf{U}^{(v)} \mathbf{U}^{(v)^T} \mathbf{U}^{(w)} \mathbf{U}^{(w)^T})}_{co-regularization\ \ term}$$

$$\text{s.t.} \quad \mathbf{U}^{(v)^T} \mathbf{U}^{(v)} = I, \ \mathbf{U}^{(w)^T} \mathbf{U}^{(w)} = I$$

- Hyperparameter λ trades off individual objectives vs co-regularization

Co-regularized Spectral Clustering

- The objective becomes:

$$\max_{\mathbf{U}^{(v)} \in \mathbb{R}^{N \times K}, \mathbf{U}^{(w)} \in \mathbb{R}^{N \times K}} \underbrace{tr(\mathbf{U}^{(v)^T} \mathcal{L}^{(v)} \mathbf{U}^{(v)})}_{view\ v\ objective} + \underbrace{tr(\mathbf{U}^{(w)^T} \mathcal{L}^{(w)} \mathbf{U}^{(w)})}_{view\ w\ objective} \underbrace{+ \lambda tr(\mathbf{U}^{(v)} \mathbf{U}^{(v)^T} \mathbf{U}^{(w)} \mathbf{U}^{(w)^T})}_{co-regularization\ term}$$

$$\text{s.t.} \quad \mathbf{U}^{(v)^T} \mathbf{U}^{(v)} = I,\ \mathbf{U}^{(w)^T} \mathbf{U}^{(w)} = I$$

- Hyperparameter λ trades off individual objectives vs co-regularization
- Can be solved using an **alternating optimization** scheme
- For a fixed $\mathbf{U}^{(w)}$, we get the following optimization problem in $\mathbf{U}^{(v)}$:

$$\max_{\mathbf{U}^{(v)} \in \mathbb{R}^{N \times K}} tr\{\mathbf{U}^{(v)^T} (\mathcal{L}^{(v)} + \lambda \mathbf{U}^{(w)} \mathbf{U}^{(w)^T}) \mathbf{U}^{(v)}\}, \quad \text{s.t.} \quad \mathbf{U}^{(v)^T} \mathbf{U}^{(v)} = I$$

Co-regularized Spectral Clustering

- The objective becomes:

$$\max_{\mathbf{U}^{(v)}\in\mathbb{R}^{N\times K},\mathbf{U}^{(w)}\in\mathbb{R}^{N\times K}} \underbrace{tr(\mathbf{U}^{(v)^T}\mathcal{L}^{(v)}\mathbf{U}^{(v)})}_{view\ v\ objective} + \underbrace{tr(\mathbf{U}^{(w)^T}\mathcal{L}^{(w)}\mathbf{U}^{(w)})}_{view\ w\ objective} \underbrace{+\lambda tr(\mathbf{U}^{(v)}\mathbf{U}^{(v)^T}\mathbf{U}^{(w)}\mathbf{U}^{(w)^T})}_{co-regularization\ term}$$

$$\text{s.t.}\quad \mathbf{U}^{(v)^T}\mathbf{U}^{(v)} = I,\ \mathbf{U}^{(w)^T}\mathbf{U}^{(w)} = I$$

- Hyperparameter λ trades off individual objectives vs co-regularization
- Can be solved using an **alternating optimization** scheme
- For a fixed $\mathbf{U}^{(w)}$, we get the following optimization problem in $\mathbf{U}^{(v)}$:

$$\max_{\mathbf{U}^{(v)}\in\mathbb{R}^{N\times K}} tr\{\mathbf{U}^{(v)^T}(\mathcal{L}^{(v)} + \lambda\mathbf{U}^{(w)}\mathbf{U}^{(w)^T})\mathbf{U}^{(v)}\}, \quad \text{s.t.}\ \mathbf{U}^{(v)^T}\mathbf{U}^{(v)} = I$$

- Equivalent to standard spectral clustering with a **modified Laplacian**

$$\mathcal{L}^{(v)} \rightarrow \mathcal{L}^{(v)} + \lambda\mathbf{U}^{(w)}\mathbf{U}^{(w)^T}$$ (Akin to Kernel or Laplacian combination)

Co-regularized Spectral Clustering

- The objective becomes:

$$\max_{\mathbf{U}^{(v)} \in \mathbb{R}^{N\times K}, \mathbf{U}^{(w)} \in \mathbb{R}^{N\times K}} \underbrace{tr(\mathbf{U}^{(v)^T} \mathcal{L}^{(v)} \mathbf{U}^{(v)})}_{view\ v\ objective} + \underbrace{tr(\mathbf{U}^{(w)^T} \mathcal{L}^{(w)} \mathbf{U}^{(w)})}_{view\ w\ objective} \underbrace{+\lambda tr(\mathbf{U}^{(v)}\mathbf{U}^{(v)^T}\mathbf{U}^{(w)}\mathbf{U}^{(w)^T})}_{co-regularization\ term}$$

$$\text{s.t.} \quad \mathbf{U}^{(v)^T}\mathbf{U}^{(v)} = I, \ \mathbf{U}^{(w)^T}\mathbf{U}^{(w)} = I$$

- Hyperparameter λ trades off individual objectives vs co-regularization
- Can be solved using an **alternating optimization** scheme
- For a fixed $\mathbf{U}^{(w)}$, we get the following optimization problem in $\mathbf{U}^{(v)}$:

$$\max_{\mathbf{U}^{(v)} \in \mathbb{R}^{N\times K}} tr\{\mathbf{U}^{(v)^T}(\mathcal{L}^{(v)} + \lambda \mathbf{U}^{(w)}\mathbf{U}^{(w)^T})\mathbf{U}^{(v)}\}, \quad \text{s.t.} \ \mathbf{U}^{(v)^T}\mathbf{U}^{(v)} = I$$

- Equivalent to standard spectral clustering with a **modified Laplacian**

$$\mathcal{L}^{(v)} \rightarrow \mathcal{L}^{(v)} + \lambda \mathbf{U}^{(w)}\mathbf{U}^{(w)^T}$$ (Akin to Kernel or Laplacian combination)

- Iteratively (in alternating fashion) solve for $\mathbf{U}^{(v)}$ and $\mathbf{U}^{(w)}$ until convergence
- Guaranteed to converge

Experiments

- Comparisons against a number of baselines. Prominent ones:
  - CCA: Clustering with features extracted by combining multiple views
  - Minimization-Disagreement: A multiview spectral clustering algorithm (de Sa, Machine Learning Journal, 2010)
- Results on **UCI Handwritten digits data** (view 1: Fourier coefficients, view 2: profile correlations)

Method	F-score	Precision	Recall	Entropy	NMI	Adj-RI
Best Single View	0.577(0.015)	0.569(0.020)	0.586(0.012)	1.198(0.029)	0.641(0.008)	0.530(0.017)
Feature Concat	0.536(0.027)	0.514(0.026)	0.561(0.032)	1.283(0.050)	0.619(0.015)	0.480(0.026)
Kernel Addition	0.707(0.052)	0.688(0.065)	0.727(0.037)	0.862(0.110)	0.744(0.030)	0.673(0.059)
Kernel Product	0.719(0.049)	0.698(0.064)	0.742(0.032)	0.832(0.102)	0.754(0.026)	0.687(0.055)
CCA	0.638(0.027)	0.616(0.037)	0.662(0.020)	1.073(0.071)	0.682(0.019)	0.596(0.031)
Min-Disagreement	0.693(0.047)	0.663(0.066)	0.729(0.026)	0.870(0.096)	0.745(0.024)	0.658(0.053)
Co-regularized	0.725(0.053)	0.707(0.067)	0.745(0.037)	0.813(0.116)	0.759(0.031)	0.694(0.060)

Experiments

- Comparisons against a number of baselines. Prominent ones:
  - CCA: Clustering with features extracted by combining multiple views
  - Minimization-Disagreement: A multiview spectral clustering algorithm (de Sa, Machine Learning Journal, 2010)
- Results on **UCI Handwritten digits data** (view 1: Fourier coefficients, view 2: profile correlations)

Method	F-score	Precision	Recall	Entropy	NMI	Adj-RI
Best Single View	0.577(0.015)	0.569(0.020)	0.586(0.012)	1.198(0.029)	0.641(0.008)	0.530(0.017)
Feature Concat	0.536(0.027)	0.514(0.026)	0.561(0.032)	1.283(0.050)	0.619(0.015)	0.480(0.026)
Kernel Addition	0.707(0.052)	0.688(0.065)	0.727(0.037)	0.862(0.110)	0.744(0.030)	0.673(0.059)
Kernel Product	0.719(0.049)	0.698(0.064)	0.742(0.032)	0.832(0.102)	0.754(0.026)	0.687(0.055)
CCA	0.638(0.027)	0.616(0.037)	0.662(0.020)	1.073(0.071)	0.682(0.019)	0.596(0.031)
Min-Disagreement	0.693(0.047)	0.663(0.066)	0.729(0.026)	0.870(0.096)	0.745(0.024)	0.658(0.053)
Co-regularized	0.725(0.053)	0.707(0.067)	0.745(0.037)	0.813(0.116)	0.759(0.031)	0.694(0.060)

- Results on **Reuters multilingual data** (view 1: English, view 2: French)

Method	F-score	Precision	Recall	Entropy	NMI	Adj-RI
Best Single View	0.342(0.010)	0.296(0.015)	0.407(0.025)	1.878(0.052)	0.287(0.019)	0.186(0.014)
Feature Concat	0.368(0.012)	0.330(0.016)	0.416(0.017)	1.841(0.057)	0.298(0.020)	0.225(0.017)
Kernel Addition	0.386(0.012)	0.358(0.017)	0.420(0.023)	1.770(0.058)	0.323(0.021)	0.252(0.016)
Kernel Product	0.258(0.003)	0.198(0.011)	0.381(0.058)	2.306(0.034)	0.123(0.010)	0.052(0.014)
CCA	0.262(0.007)	0.222(0.005)	0.322(0.034)	2.232(0.009)	0.147(0.003)	0.082(0.003)
Min-Disagreement	0.381(0.014)	0.341(0.004)	0.435(0.035)	1.736(0.052)	0.342(0.024)	0.240(0.012)
Co-regularized	0.405(0.001)	0.357(0.003)	0.467(0.011)	1.654(0.003)	0.375(0.002)	0.267(0.001)

- Sensitivity to the co-regularization parameter λ
  - Better than the closest performing baseline for a wide range of λ
- Clustering performance vs iterations
  - Performance stabilizes within very few iterations

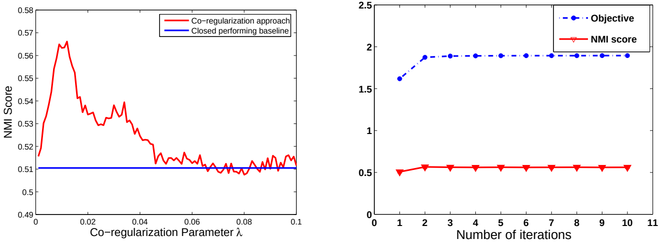


Figure: On Caltech-101 data. (Left) Effect of varying the co-regularization parameter. (Right) Clustering performance vs iterations

Conclusion

- Proposed a co-regularization approach to an unsupervised learning problem
- Can be seen as combining multiple kernels
- Objective leads to a simple eigenvalue problem
  - Can be efficiently solved by state-of-the-art eigensolvers
- Regularizers other than $tr(\mathbf{K}_{\mathbf{U}^{(v)}}\mathbf{K}_{\mathbf{U}^{(w)}})$ could also be tried
- Can be applied to solve other unsupervised multiview learning problems
  - E.g., spectral methods for dimensionality reduction (Kernel PCA)

Conclusion

- Proposed a co-regularization approach to an unsupervised learning problem
- Can be seen as combining multiple kernels
- Objective leads to a simple eigenvalue problem
  - Can be efficiently solved by state-of-the-art eigensolvers
- Regularizers other than $tr(\mathbf{K}_{\mathbf{U}^{(v)}}\mathbf{K}_{\mathbf{U}^{(w)}})$ could also be tried
- Can be applied to solve other unsupervised multiview learning problems
  - E.g., spectral methods for dimensionality reduction (Kernel PCA)

Thanks! Questions?