Spectral Clustering with Inconsistent Advice

T. Coleman  J. Saunderson  A. Wirth

Department of Computer Science
University of Melbourne

July 7, 2008
Outline

1. The Problem
   - Clustering Problems
   - Advice
   - Clustering with Inconsistent Advice

2. The Solution
   - Spectral Clustering
   - From 2CC to subspaces
   - Finding families of low-cost 2CC solutions

3. Experiments
Clustering is a tool for exploring data.
Clustering

- Clustering is a tool for exploring data.
- We want to form groups of similar datapoints.
Clustering is a tool for exploring data.

We want to form groups of similar datapoints.
Clustering is a tool for exploring data.

We want to form groups of similar datapoints.
Clustering is a tool for exploring data.

We want to form groups of *similar* datapoints.
Clustering

• Clustering is a tool for exploring data.
• We want to form groups of similar datapoints.
We assume we only have pairwise “affinity” information:
We assume we only have pairwise “affinity” information:
Input: Affinities

- We assume we only have pairwise “affinity” information:
We assume we only have pairwise “affinity” information:
We assume we only have pairwise “affinity” information:

So input is matrix:

\[ A = \begin{bmatrix} \cdots & a_{ij} \\ \vdots & \ddots \end{bmatrix} \]
Clusterining Costs

- How do we rate a clustering? (so we can find the best)
Clustering Costs

- How do we rate a clustering? (so we can find the best)
- All we know is affinities (similarities) between points.
Clustering Costs

- How do we rate a clustering? (so we can find the best)
- All we know is affinities (similarities) between points.
- Cut cost:

$$\text{CUT}(C_1, C_2) = \sum_{x \in C_1, y \in C_2} A(x, y)$$
How do we rate a clustering? (so we can find the best)
- All we know is affinities (similarities) between points.
- Cut cost:
  \[
  \text{CUT}(C_1, C_2) = \sum_{x \in C_1, y \in C_2} A(x, y)
  \]
- Problem: The best clustering is \((V, \emptyset)\)
Clustering Costs

- How do we rate a clustering? (so we can find the best)
- All we know is affinities (similarities) between points.
- Cut cost:
  \[ \text{CUT}(C_1, C_2) = \sum_{x \in C_1, y \in C_2} A(x, y) \]
- Problem: The best clustering is \((V, \emptyset)\)
- Ratio-cut cost:
  \[ \text{RCUT}(C_1, C_2) = \frac{\text{CUT}}{|C_1||C_2|} \]
Clustering Costs

- How do we rate a clustering? (so we can find the best)
- All we know is affinities (similarities) between points.
- Cut cost:
  \[ \text{CUT}(C_1, C_2) = \sum_{x \in C_1, y \in C_2} A(x, y) \]
- Problem: The best clustering is \((V, \emptyset)\)
- Ratio-cut cost:
  \[ \text{RCUT}(C_1, C_2) = \frac{\text{CUT}}{|C_1||C_2|} \]
- Normalized-cut cost:
  \[ \text{NCUT}(C_1, C_2) = \frac{\text{CUT}}{\text{vol}(C_1)\text{vol}(C_2)} \]
Advice for Clustering Problems

- We may know more information that just affinities between items.
We may know more information that just affinities between items.

We might have *advice* about which items should be in the same cluster, and which should not be in the same cluster.
We may know more information that just affinities between items.

We might have *advice* about which items should be in the same cluster, and which should not be in the same cluster.

However, this advice may not be absolute (i.e. constraints).
Advice for Clustering Problems

- We may know more information that just affinities between items.
- We might have *advice* about which items should be in the same cluster, and which should not be in the same cluster.
- However, this advice may not be absolute (i.e. constraints).
- Also, it may not be consistent.
Advice for Clustering Problems

- We may know more information that just affinities between items.
- We might have advice about which items should be in the same cluster, and which should not be in the same cluster.
- However, this advice may not be absolute (i.e. constraints).
- Also, it may not be consistent.
The 2-Correlation Clustering Problem

- If we are only interested in the advice, we will be solving the two-correlation-clustering (2CC) problem.
If we are only interested in the advice, we will be solving the two-correlation-clustering (2CC) problem.
The 2-Correlation Clustering Problem

If we are only interested in the advice, we will be solving the two-correlation-clustering (2CC) problem.
The 2-Correlation Clustering Problem

- If we are only interested in the advice, we will be solving the two-correlation-clustering (2CC) problem.
The 2-Correlation Clustering Problem

If we are only interested in the advice, we will be solving the two-correlation-clustering (2CC) problem.
If we are only interested in the advice, we will be solving the two-correlation-clustering (2CC) problem.
The 2-Correlation Clustering Problem

If we are only interested in the advice, we will be solving the two-correlation-clustering (2CC) problem.

2CC asks us to minimize:

\[ \text{Number of } + \text{ edges between clusters } + \]
\[ \text{Number of } - \text{ edges within clusters } \]
We have two problems to simultaneously solve:
We have two problems to simultaneously solve:

- **Rcut**
- **2CC**
Outline

1. The Problem
   - Clustering Problems
   - Advice
   - Clustering with Inconsistent Advice

2. The Solution
   - Spectral Clustering
   - From 2CC to subspaces
   - Finding families of low-cost 2CC solutions

3. Experiments
Ratio Cut

Recall:

$$RCUT(C_1, C_2) = \frac{\sum_{x \in C_1, y \in C_2} A(x, y)}{|C_1||C_2|}$$
Ratio Cut

Recall:

\[ R_{\text{cut}}(C_1, C_2) = \frac{\sum_{x \in C_1, y \in C_2} A(x, y)}{|C_1||C_2|} \]

This can be rewritten as

\[ R_{\text{cut}}(v) = \frac{\sum_{i < j} a_{ij} \|v_i - v_j\|^2}{\sum_{i < j} \|v_i - v_j\|^2} \]

where

- \( v \in \{-1, 1\}^n \)
Recall:
\[
R_{\text{cut}}(C_1, C_2) = \frac{\sum_{x \in C_1, y \in C_2} A(x, y)}{|C_1||C_2|}
\]

This can be rewritten as
\[
R_{\text{cut}}(v) = \frac{\sum_{i<j} a_{ij} \|v_i - v_j\|^2}{\sum_{i<j} \|v_i - v_j\|^2}
\]

where
- \( v \in \{-1, 1\}^n \)
- We will relax so that \( v \in \mathbb{R} \)
Suppose we have a problem of the form: 

\[ \min_{v \in \mathbb{R}^n} \frac{v^T A v}{v^T v} \]
The Rayleigh-Ritz Theorem

- Suppose we have a problem of the form: \((A \succeq 0)\)

\[
\min_{\nu \in \mathbb{R}^n} \frac{\nu^T A \nu}{\nu^T \nu}
\]

- Then the solution is given by \(\nu_{\min}(A)\), where \(\nu_{\min}(X)\) is the smallest eigenvector of \(X\).
The Rayleigh-Ritz Theorem

Suppose we have a problem of the form: \((A \succeq 0)\)

\[
\min_{v \in \mathbb{R}^n} \frac{v^T A v}{v^T v}
\]

Then the solution is given by \(v_{\min}(A)\), where \(v_{\min}(X)\) is the smallest eigenvector of \(X\).

The cost is \(\lambda_{\min}, v_{\min}'s\) eigenvalue.
The Rayleigh-Ritz Theorem

- Suppose we have a problem of the form: \( (A \succeq 0) \)

\[
\min_{\mathbf{v} \in \mathbb{R}^n} \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}}
\]

- Then the solution is given by \( \mathbf{v}_{\min}(A) \), where \( \mathbf{v}_{\min}(X) \) is the smallest eigenvector of \( X \).
- The cost is \( \lambda_{\min} \), \( \mathbf{v}_{\min} \)'s eigenvalue.

Extended Rayleigh-Ritz:

- If further, we constrain \( \mathbf{v} \in S \), then:

\[
Y \mathbf{v}_{\min}(Y^T A Y)
\]

- \( Y \) is a matrix of orthonormal basis vectors for \( S \).
Spectral Clustering Solution

Spectral clustering:

\[
\min_{\nu} \text{RCUT}(\nu) \quad \text{such that} \quad \nu \in \text{‘non-trivial subspace’}
\]

- The solution is given by \( \nu = \text{eigenvector of } L(A) \) corresponding to the \textit{second smallest} eigenvlalue.

- A significant attraction of spectral clustering is that there are well-developed numerical algorithms for solving eigenproblems.
Spectral Clustering with Subspace Constraints

More general problem:

$$\min_{v} \text{Rcut}(v) \quad \text{such that}$$

$$v \in \text{‘non-trivial subspace’ } \cap S$$

where $S$ is any subspace of $\mathbb{R}^n$. 
Spectral Clustering with Subspace Constraints

More general problem:

$$\min_{\nu} \text{RCUT}(\nu) \quad \text{such that}$$

$$\nu \in \text{‘non-trivial subspace’} \cap S$$

where $S$ is any subspace of $\mathbb{R}^n$.

- The solution is also given by an eigenvector of some matrix.
- Essentially the solution is the best spectral clustering in the allowed subspace.
Spectral Clustering with Subspace Constraints

More general problem:

\[ \min_{\nu} \text{RCUT}(\nu) \quad \text{such that} \quad \nu \in \text{‘non-trivial subspace’} \cap S \]

where \( S \) is any subspace of \( \mathbb{R}^n \).

- The solution is also given by an eigenvector of some matrix.
- Essentially the solution is the best spectral clustering \textit{in the allowed subspace}.

**Main idea**

Use some method to transform the advice into a subspace \( S \), and solve spectral clustering on this subspace.
The ‘subspace trick’ [De Bie, Suykens, De Moor (2004)]

- This is a method for converting each piece of advice into a subspace.
The ‘subspace trick’ [De Bie, Suykens, De Moor (2004)]

- This is a method for converting each piece of advice into a subspace.

Advice:
- \( v_1 \) must be in same cluster as \( v_2 \)
- \( v_1 \) must be in different cluster to \( v_2 \)

Subspace:
- \( \{ v \in \mathbb{R}^n : v_1 = v_2 \} \)
- \( \{ v \in \mathbb{R}^n : v_1 = -v_2 \} \)
The ‘subspace trick’ [De Bie, Suykens, De Moor (2004)]

This is a method for converting each piece of advice into a subspace.

Advice:
\[ v_1 \text{ must be in same cluster as } v_2 \]
\[ v_1 \text{ must be in different cluster to } v_2 \]

Subspace:
\[ \{ v \in \mathbb{R}^n : v_1 = v_2 \} \]
\[ \{ v \in \mathbb{R}^n : v_1 = -v_2 \} \]

Take \( S \) to be the intersection (over all pieces of advice) of these subspaces (call them single-advice-subspaces).
An example

Advice: $v_1 = v_2$
Method One

- If the advice is inconsistent then when we take the intersection of the single-advice-subspaces we get \( \{0\}! \)
Method One

If the advice is inconsistent then when we take the intersection of the single-advice-subspaces we get \{0\}!

Method One

1. (Somehow) solve 2CC.
2. Let $S$ be the intersection of the single-advice-subspaces corresponding to the advice that the 2CC solution does not violate.
Method One

1. (Somehow) solve 2CC.
2. Let $S$ be the intersection of the single-advice-subspaces corresponding to the advice that the 2CC solution does not violate.
**Method One**

1. (Somehow) solve 2CC.
2. Let \( S \) be the intersection of the single-advice-subspaces corresponding to the advice that the 2CC solution does *not* violate.
Method One

1. (Somehow) solve 2CC.
2. Let $S$ be the intersection of the single-advice-subspaces corresponding to the advice that the 2CC solution does not violate.

Problem: the combinatorial solution of 2CC that we use may not be unique:
- Choosing one such solution over another may be unwise.
## Method Two

1. (Somehow) find *all* solutions of 2CC.
2. Each of these can be converted into subspaces $S_1, S_2, \ldots, S_m$ as in Method One.
3. Let $S$ be the smallest subspace containing $S_1, S_2, \ldots, S_m$. 
Method Two

1. (Somehow) find all solutions of 2CC.
2. Each of these can be converted into subspaces $S_1, S_2, \ldots, S_m$ as in Method One.
3. Let $S$ be the smallest subspace containing $S_1, S_2, \ldots, S_m$.

Potential problem: perhaps we are being too respectful of the advice. After all, it could be very noisy.
Method Three

1. (Somehow) find all clusterings that have 2CC cost at most $f$ times bigger than the optimal cost.

2. Each of these can be converted into subspaces and combined as in Method Two.

But... we can’t even solve 2CC, let alone find all the possible solutions and all the ‘near’-solutions!
Spectral relaxation of 2CC

Like $R_{\text{cut}}$ we can write 2CC as

$$2CC(v) = \frac{\sum_{i<j} \|v_i - w_{ij}v_j\|^2}{\|v\|^2}$$

where $v \in \{-1, 1\}^n$ and $w_{ij}$ is the label on edge $(ij)$
Like $\text{RCUT}$ we can write 2CC as

$$2CC(v) = \frac{\sum_{i<j} \|v_i - w_{ij}v_j\|^2}{\|v\|^2}$$

where $v \in \{-1, 1\}^n$ and $w_{ij}$ is the label on edge $(ij)$

- Again we relax to allow $v \in \mathbb{R}^n$. 

Like $R_{\text{CUT}}$ we can write 2CC as

$$2CC(v) = \frac{\sum_{i<j} \|v_i - w_{ij}v_j\|^2}{\|v\|^2}$$

where $v \in \{-1, 1\}^n$ and $w_{ij}$ is the label on edge $(ij)$

- Again we relax to allow $v \in \mathbb{R}^n$.
- The solution is the eigenspace corresponding to smallest eigenvalue of a matrix associated with the $w_{ij}$.
Like $\text{RCUT}$ we can write 2CC as

$$2CC(v) = \frac{\sum_{i<j} \| v_i - w_{ij} v_j \|^2}{\| v \|^2}$$

where $v \in \{-1, 1\}^n$ and $w_{ij}$ is the label on edge $(ij)$.

- Again we relax to allow $v \in \mathbb{R}^n$.
- The solution is the eigenspace corresponding to smallest eigenvalue of a matrix associated with the $w_{ij}$.
- Cost of the solution is (proportional to) the corresponding eigenvalue!
A natural family of subspaces

- For **Method Three** we wanted to produce a family of subspaces containing 2CC solutions of cost at most $f$ times optimum.

- Each of the eigenspaces $S_\lambda$ with eigenvalue at most $f$ times $\lambda_{min}$ contains 2CC solutions of *relaxed* cost at most $f$ times the optimum *relaxed* cost.
For Method Three we wanted to produce a family of subspaces containing 2CC solutions of cost at most $f$ times optimum.

Each of the eigenspaces $S_\lambda$ with eigenvalue at most $f$ times $\lambda_{min}$ contains 2CC solutions of relaxed cost at most $f$ times the optimum relaxed cost.

Method Three (relaxed)

1. Find eigenspaces $S_\lambda$ corresponding to eigenvalues at most $f$ times $\lambda_{min}$.
2. Let $S$ be the smallest subspace of $\mathbb{R}^n$ containing all such $S_\lambda$. 
Summary of solution method

- Choose $f$ factor.
- Use relaxed version of Method Three to convert the advice into a subspace $S$.
- Solve spectral clustering constrained to $S$ to give $\nu \in \mathbb{R}^n$.
- Round to give $\nu \in \{-1, 1\}^n$ i.e. a clustering.
The Problem
- Clustering Problems
- Advice
- Clustering with Inconsistent Advice

The Solution
- Spectral Clustering
- From 2CC to subspaces
- Finding families of low-cost 2CC solutions

Experiments
Datasets

- We tested our algorithms on standard UCI classification datasets.
Datasets

- We tested our algorithms on standard UCI classification datasets.
- Class labels unknown, cluster sizes not uniform.
Datasets

- We tested our algorithms on standard UCI classification datasets.
- Class labels unknown, cluster sizes not uniform.
- Advice data generated probabilistically.
Datasets

- We tested our algorithms on standard UCI classification datasets.
- Class labels unknown, cluster sizes not uniform.
- Advice data generated probabilistically.
  - Complete advice—solvable only heuristically.
Datasets

- We tested our algorithms on standard UCI classification datasets.
- Class labels unknown, cluster sizes not uniform.
- Advice data generated probabilistically.
  - Complete advice—solvable only heuristically.
  - Dense advice—SDP solvable.
Heart Disease, Dense Advice, p=0.75

The Problem
The Solution
Experiments

Accuracy

f value

T. Coleman, J. Saunderson, A. Wirth
Spectral Clustering with Inconsistent Advice
There are problems where a combination of advice and spectral clustering can improve on each technique separately.
Conclusions

- There are problems where a combination of advice and spectral clustering can improve on each technique separately.
- We’ve demonstrated an approach and one algorithm to use.
Conclusions

- There are problems where a combination of advice and spectral clustering can improve on each technique separately.
- We’ve demonstrated an approach and one algorithm to use.
- Further Work
There are problems where a combination of advice and spectral clustering can improve on each technique separately.

We’ve demonstrated an approach and one algorithm to use.

**Further Work**
- Other algorithms to fit the framework
Conclusions

- There are problems where a combination of advice and spectral clustering can improve on each technique separately.
- We’ve demonstrated an approach and one algorithm to use.

**Further Work**

- Other algorithms to fit the framework
- It is not obvious how to choose the parameters
There are problems where a combination of advice and spectral clustering can improve on each technique separately.

We’ve demonstrated an approach and one algorithm to use.

**Further Work**

- Other algorithms to fit the framework
- It is not obvious how to choose the parameters
- An extension to $k$ clusters?
There are problems where a combination of advice and spectral clustering can improve on each technique separately.

We’ve demonstrated an approach and one algorithm to use.

Further Work
- Other algorithms to fit the framework
- It is not obvious how to choose the parameters
- An extension to $k$ clusters?

Acknowledgements
- Ian Davidson for suggesting the problem
- Anonymous reviewers for helpful comments
- ARC Discovery Grant DP0663979