University of Turin, Italy
Department of Computer Science

Parameter-free Hierarchical Co-Clustering by *n*-Ary Splits

Dino Ienco, Ruggero G. Pensa and Rosa Meo
{ienco,pensa,meo}@di.unito.it

ECML-PKDD 2009 – Bled (Slovenia)

Outline

Motivations

Our Idea

Co-Clustering and Background

Hierachical Co-Clustering

Results

Conclusions

Motivations

Co-Clustering:

- effective approach that obtains interesting results
- Commonly involved with high-dimensional data
- Partition simultaneously rows and columns

Many Co-clustering algorithms:

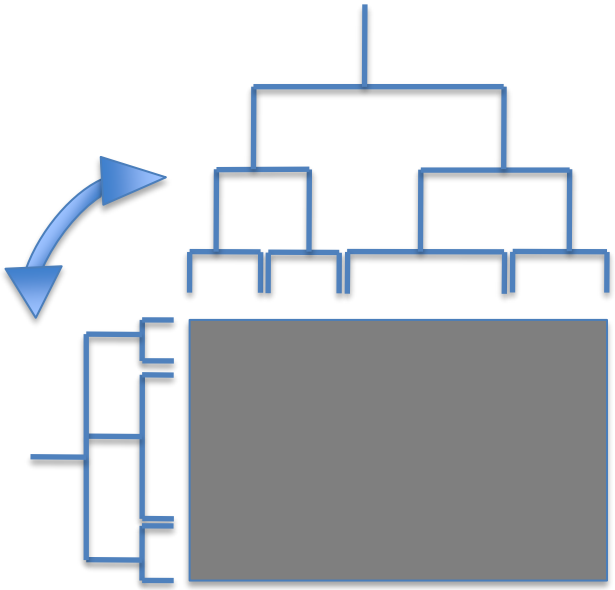
- Spectral approach (Dhillon et al. KDD01)
- Information theoretic approach (Dhillon et al. KDD03)
- Minimum Sum-Squared Residue approach (Cho et al. SDM04)
- Bayesian approach (Shan et al. ICDM08)

- *All previous techniques*:
  - require num. of row/column cluster as parameter
  - produce flat partitions, without any structure information

In general:

- parameters are difficult to set
- structured output (like hierarchies)

help the user to understand data



Hierarchical structures are useful to:

- indexing and visualize data
- explore the parent-child relationships
- derive generalization/specialization concept

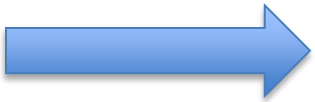
Our Idea

CO-CLUSTERING

+

HIERARCHICAL APPROACH

ALLOWS



Build two hierarchies on both dimensions

SIMULTANEOUSLY

PROPOSED APPROACH:

- Extend previous flat co-clustering algorithm (Robardet02)

to hierarchical setting

Background

τ-CoClust *(Robardet02):*

- Co-Clustering for **counting** or **frequency data**
- **No number of row/column** clustering needed
- Maximize a statistical measure **Goodman and Kruskal τ** between row and column partitions

Goodman and Kruskal τ :

- Measure the proportional reduction in the prediction error of a dep. Variable given an indep. Variable

	F_1	**F_2**	**F_3**
O_1	d_{11}	d_{12}	d_{13}
O_2	d_{21}	d_{22}	d_{23}
O_3	d_{31}	d_{32}	d_{33}
O_4	d_{41}	d_{42}	d_{43}

$$t_{ij} = \sum_{O_x \in CO_k} \sum_{F_y \in CF_l} d_{xy}$$

$CO_1 = \{O_1, O_2\}$
$CO_2 = \{O_3, O_4\}$

$CF_1 = \{F_2\}$
$CF_2 = \{F_1, F_3\}$

	CF_1	**CF_2**	
CO_1	t_{11}	t_{12}	TO_1
CO_2	t_{21}	t_{22}	TO_2
	TF_1	TF_2	

Goodman and Kruskal τ :

- Measure the proportional reduction in the prediction error of a dep. Variable given an indep. Variable

$\mathbf{E_{CO}}$ Prediction error on CO without knowledge about CF partition

$\mathbf{E_{CO|CF}}$ Prediction error on CO with knowledge about CF partition

$$\tau_{CO|CF} = \frac{E_{CO} - E_{CO|CF}}{E_{CO}}$$

Optimization strategy:

- τ is asymmetrical, for this reason the algorithm alternates the optimization of two functions $\tau_{CO|CF}$ and $\tau_{CF|CO}$

- Stochastic optimization (example on rows):

```
# Start with an initial parition on rows
for i in 1..n_times
    # augment the current partition with an empty cluster
    # Move at random one element from a partition to another one
    # If obj. func. improve keep solution, else undo the operation
    # If there is an empty cluster, remove it
end
```

- This optimization allows the num. of clusters to grow or decrease

- In (Robardet02) an efficient way to update incrementally the objective function was introduced

HIERARCHICAL CO-CLUSTERING

HiCC:

- **Hierarchical Co-Clustering** algorithm that extends τ-CoClust

- **Divisive Approach**

- **No parameter settings** needed

- **No predefined number of splits** for each node of the hierarchy

HiCC:

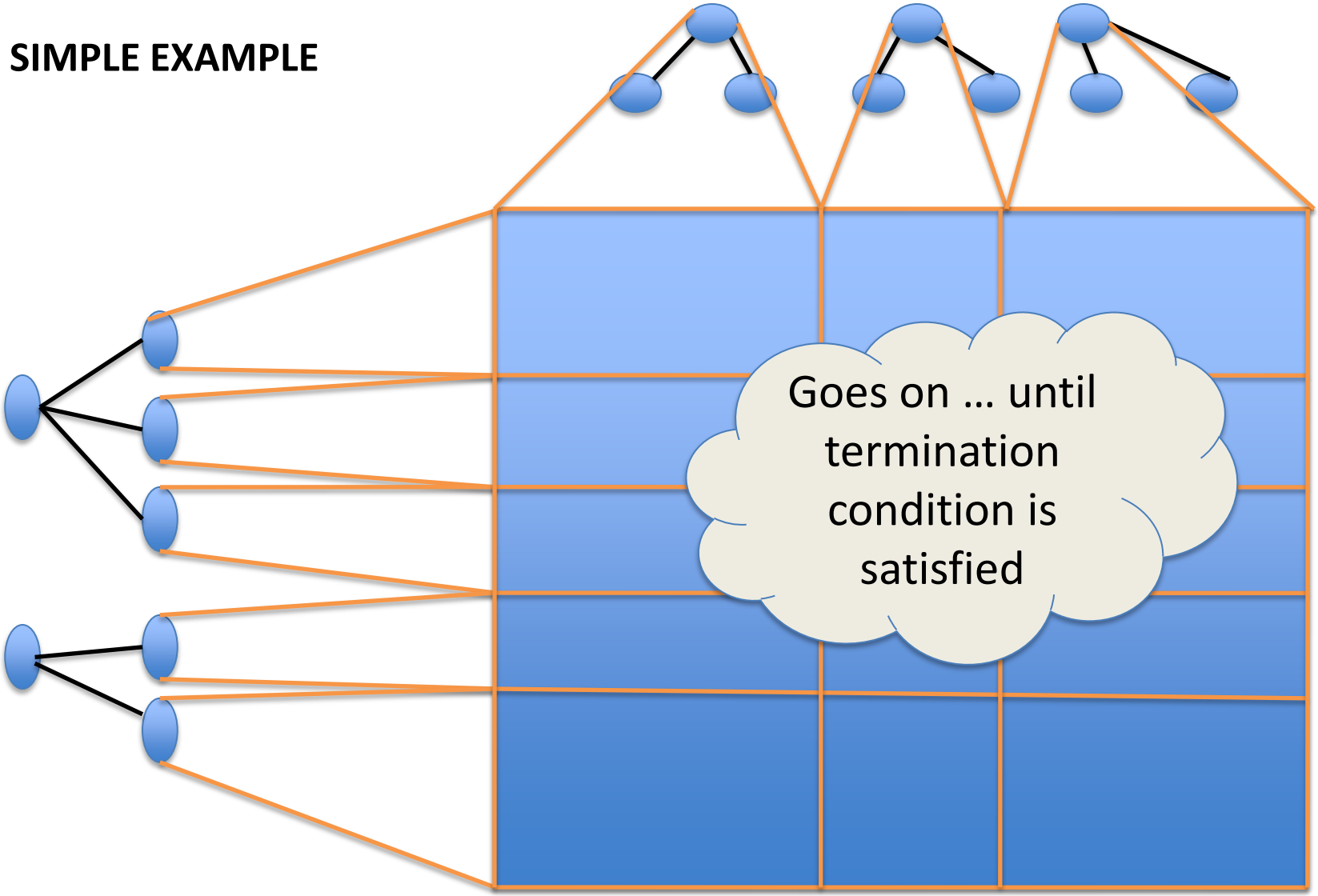
At the beginning use τ-CoClust

repeat

- **From the current Row/Column partitions**
- **Fix the Column partition**
- **For each cluster in the Row partition Re-cluster** with τ-CoClust and optimize the obj. func. $\tau_{CO|CF}$
- **Update Row Hierarchy**
- **Fix the new Row partition**
- **For each cluster in the Column partition Re-cluster** with τ-CoClust and optimize the obj. func. $\tau_{CF|newCO}$
- **Update Column Hierarchy**

until (TERMINATION)

A SIMPLE EXAMPLE



Experimentation:

- No previous hierarchical co-clustering algorithm exists

- Use a flat co-clustering algorithm with the same number of clusters obtained by our approach for each level

- We choose Information theoretic approach (KDD03) and for each level we perform 50 runs then we compute the average

- We use document-word dataset to validate our approach:
  * OHSUMED (collection of pubmed abstract) {oh0, oh15}
  * REUTERS-21578 (collected and labeled by Carnegie Group) {re0, re1}
  * TREC (Text Retrieval Conference) {tr11, tr21}

An example of row hierachy on OHSUMED

We label each cluster with the majority class

<table>
<tr><td rowspan="4">Enzyme-Activation</td><td rowspan="2">Enzyme-Activation</td><td>Enzyme-Activation</td></tr>
<tr><td>Enzyme-Activation</td></tr>
<tr><td rowspan="2">Cell-Movement</td><td>Cell-Movement</td></tr>
<tr><td>Adenosine-Diphosphate</td></tr>
<tr><td rowspan="4">Staphylococcal-Infection</td><td rowspan="2">Uremia</td><td>Uremia</td></tr>
<tr><td>Staphylococcal-Infection</td></tr>
<tr><td rowspan="2">Staphylococcal-Infection</td><td>Staphylococcal-Infection</td></tr>
<tr><td>Memory</td></tr>
</table>

An example of column hierachy on REUTERS

We label each cluster with top 10 words ranked by mutual information

<table>
<tr><td colspan="2">oil, compani, opec, gold, ga, barrel, strike, mine, lt, explor</td><td colspan="2">tonne, wheate, sugar, corn, mln, crop, grain, agricultur, usda, soybean</td><td colspan="2">coffee, buffer, cocoa, deleg, consum, ico, stock, quota, icco, produc</td></tr>
<tr><td>oil, opec, tax, price, dlr, crude, bank, industri, energi, saudi</td><td>compani, gold, mine, barrel, strike, ga, lt, ounce, ship, explor</td><td>tonne, wheate, sugar, corn, grain, crop, agricultur, usda, soybean, soviet</td><td>mln, export, farm, ec, import, market, total, sale, trader, trade</td><td>quota, stock, produc, meet, intern, talk, bag, agreem, negoti, brazil</td><td>coffee, deleg, buffer, cocoa, consum, ico, icco, pact, council, rubber</td></tr>
</table>

External Validation Indices:

- Purity
- Normalized Mutual Information (NMI)
- Adjusted Rand Index

Hierarchical setting:

We combine the hierarchical result with this formula

$$Goodness(f) = \frac{\sum_{i=1} \alpha_i * f_i}{\sum_{i=1} \alpha_i}$$

- f is one of the external validation indices

- α_i is a weight for the hierarchy level i,
  in our case α_i is equal to 1/i

Performance Results

Dataset	ITCC			HiCC		
	Goodness			Goodness		
	NMI	Purity	Adj. Rand Index	NMI	Purity	Adj. Rand Index
oh0	0.4311	0.5705	0.1890	0.4607	0.5748	0.1558
oh15	0.3238	0.4628	0.1397	0.3337	0.4710	0.1297
tr11	0.3861	0.5959	0.1526	0.3949	0.6028	0.1325
tr21	0.1526	0.7291	0.0245	0.1277	0.7332	0.0426
re0	0.2820	0.5695	0.0913	0.2175	0.5388	0.0472
re1	0.3252	0.4957	0.0832	0.3849	0.5513	0.1261

Performance Results on re1 dataset

		ITCC			HiCC		
RowClust	ColClust	NMI	Purity	Adj. Rand	NMI	Purity	Adj. Rand
3	3	0.1807	0.3282	0.1028	0.3290	0.4255	0.2055
6	6	0.2790	0.3991	0.1723	0.2914	0.4255	0.1828
15	257	0.2969	0.4357	0.1358	0.3098	0.4472	0.1555
330	1291	0.3499	0.4013	0.0021	0.3950	0.51	0.0291
812	2455	0.4857	0.5930	0.0013	0.4810	0.6530	0.0031
1293	3270	0.5525	0.8284	0.0008	0.5517	0.8461	0.0009
1575	3629	0.5864	0.9602	0.0005	0.5854	0.9638	0.0001
1646	3745	0.5944	0.9926	0.0004	0.5940	0.9952	0
1657	3757	0.5951	1	0	0.5952	1	0
1657	3758	0.5951	1	0	0.5952	1	0

Conclusions

We propose:

- New approach to hierarchical co-clustering
- Parameter free
- No apriori fixed number of splits (n-ary splits)
- Obtains good results
- Builds simultaneously hierarchies on both dimensions
- Improve co-clustering results exploration

Future works:

- Parallelize the algorithm to improve time performance

- Pushing constraints inside it to use background knowledge

- Extend the framework to manage continuous data

Any Question?

Thank you for your attention