Our starting point

Many unsolved questions about clustering

In particular:
How to evaluate clusterings and compare clustering algorithms?

Imagine to organize a clustering competition

Need to figure out:
• What tasks would you choose?
• How would you evaluate the results?
• On what data sets?

Idea: ask the NIPS community 😊
Workshop: three “tracks”

1. **From Theory to Practice**: What kind of advice could theoreticians give to practitioners?

2. **From Practice to Theory**: What would practitioners like theoreticians to investigate?

3. **From Art to Science?** Is there any way to make clustering less of an art and more of a science?
Clustering: Science or Art?

Isabelle Guyon, Robert C. Williamson, Ulrike von Luxburg

December 2009
Different angles on clustering research

Ule (theoretician) asks Isabelle (practitioner):
What would I need to do to convince you to use my clustering algorithm?

Isabelle:
Actually, what should the practitioner expect? I have K-means and average linkage hierarchical clustering in my tool box. If I use something else, what could be better? I don’t believe an algorithm can find ”my” clusters.

Ule (insisting): But how could I convince you?

Isabelle: You can’t ...
Different angles on clustering research (2)

“The theoreticians of numerical taxonomy have enjoyed themselves immensely over the past decade. The mushrooming literature is quite fascinating and new developments tumble after each other. Anyone who is prepared to learn quite a deal of matrix algebra, some classical mathematical statistics, some advanced geometry, a little set theory, perhaps a little information theory and graph theory, and some computer technique, and who has access to a good computer and enjoys mathematics (as he must if he gets this far!) will probably find the development of new taximetric methods much more rewarding, more up-to-date, more ‘general’, and hence more prestigious than merely classifying plants or animals or working out their phylogenies.”

What is missing in scientific clustering literature?

Clustering is not domain-independent!

Clustering different kinds of crop:
- Biologist: wants to know genetic relations
- Farmer: wants to know how resistant they are to drought

If we ignore these differences we can only solve parts of what end-users need.
What is missing in scientific clustering literature?

(2)

We believe: most prominent “gap” in current clustering research is the question of evaluation.

- scores such as Ncut, RAND, etc. don’t tell anything about the usefulness of a clustering for a given application

- as opposed to supervised classification, no general utility function (“loss function”) for clustering

- a utility function cannot be defined if we don’t know why we need to cluster in the first place
Evaluating the usefulness of a clustering

Current practice:

- Cluster artificial data sets (mixture of Gaussians)
  - Baseline test. Does not tell anything about the usefulness.
- Evaluation of classification error on UCI data sets.
  - Class labels might not follow cluster structure
  - There might be several good clusterings
  - Not even suitable as baseline test.
- Clustering “makes sense” on a real world data set.
  - Does not really tell much:

  Farris (1981): “One must wonder what value might be attributed to a method chosen primarily for its failure to contradict preconceptions”
Evaluating the usefulness of a clustering (2)

Evaluation needs to care about how useful a given clustering is for a given purpose.

**Example 1:** Clustering as a preprocessing step for object detection

- Final goal can be quantified (error rate)
- Clustering is “just a parameter”
- Evaluate indirectly by error rate

Here it does not help at all to compute scores like

- within-cluster similarity
- stability
Evaluating the usefulness of a clustering (3)

Example 2: Clustering for exploratory data analysis.

- Goal: “present the data to the analyst such that he can see patterns and formulate interesting hypotheses about the data” (Good, 1983)
- The evaluation must involve users (like in user interface studies)!
A list of problems and their evaluations

<table>
<thead>
<tr>
<th>Task</th>
<th>Goal</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-processing</td>
<td>Prediction</td>
<td>Improved prediction performance</td>
</tr>
<tr>
<td>Compression</td>
<td>Reduce size</td>
<td>Compression rate</td>
</tr>
<tr>
<td>Exploratory data analysis</td>
<td>Scientific discovery</td>
<td>Statistical tests</td>
</tr>
<tr>
<td>Define categories</td>
<td>Taxonomy (debatable)</td>
<td>Stability across data representation</td>
</tr>
<tr>
<td>Data sanity check</td>
<td>Catch experimental artifacts</td>
<td>Statistical tests</td>
</tr>
</tbody>
</table>
More systematic approach

“Difficulties in identifying problems have delayed statistics far more than difficulties in solving problems. Thus it is appropriate to be as systematic as we can about unsolved problems. Different ends require different means and different logical structures.”


We suggest:

- Build a catalog of clustering problems
- Independent of particular algorithms
- Devise evaluation procedures for groups of problems.
More systematic approach (2)

Identify important “dimensions” to describe clustering problems:

- Qualitative — Quantitative
- Exploratory — Confirmatory
- Unsupervised — supervised
- Inductive bias
  - compact clusters — chain-like clusters
  - peak-based — gap-based
  - flat — hierarchical
- Meaningful categories — useful categories

Try to devise evaluation procedures for groups of problems.
Summary: From Art to Science

We believe:

▶ The most important question in clustering is not to come up with new objective functions or algorithms.
▶ It is pointless to compare different clustering paradigms independently of applications ("my algorithm is better than yours")
▶ But it is very important to compare algorithms, techniques etc for specific applications, evaluated by end-use criteria.

We conjecture:

▶ Helpful step: taxonomy of clustering problems