PREFLIB and Empirical Testing in Computational Social Choice

Nicholas Mattei
NICTA and UNSW
PrefLib: A Library for Preferences

• Many research communities have libraries, datasets, and tool chains that are standard and widely used.

• Preference handling and computational social choice have largely centered around theoretical results.

• We have collected datasets and tools to establish PrefLib, a library of preference data, as a service to the wider community.
Road Map

1. PrefLib: What it is and why we made it.

2. A detailed look at the data within PrefLib.

3. An overview of PrefLib: Tools and what they can do for you.
Preference Aggregation

- Problems arise when groups of agents (humans and/or computers) need to make a collective decision.

- How do we aggregate individual (possibly conflicting) preference profiles into a collective preference profile?
Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 239 data sets as a service to the machine learning community. You may view all data sets through our searchable interface. Our old web site is still available, for those who prefer the old format. For a general overview of the Repository, please visit our About page. For information about citing data sets in publications, please read our citation policy. If you wish to donate a data set, please consult our donation policy. For any other questions, feel free to contact the Repository librarians. We have also set up a mirror site for the Repository.

Latest News:
- 2013-04-04: Welcome to the new Repository admin, Kevin Bache and Moshe Lichman!
- 2010-03-01: Note from donor regarding Netflix data
- 2009-10-16: Two new data sets have been added.
- 2009-09-14: Several data sets have been added.
- 2008-07-23: Repository mirror has been set up.
- 2008-03-24: New data sets have been added!
- 2007-06-25: Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

Featured Data Set: Internet Advertisements

<table>
<thead>
<tr>
<th>Task: Classification</th>
<th>Data Type: Multivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td># Attributes: 1566</td>
<td># Instances: 3279</td>
</tr>
</tbody>
</table>

This dataset represents a set of possible advertisements on Internet pages.

Newest Data Sets:
- **2013-03-07:** UCI Daphnet Freezing of Gait
- **2013-01-17:** UCI Fertility
- **2013-01-03:** UCI Yacht Hydrodynamics
- **2012-12-10:** UCI Human Activity Recognition Using Smartphones
- **2012-12-03:** UCI One-hundred plant species leaves data set
- **2012-11-30:** UCI Energy efficiency
- **2012-10-21:** UCI Qt40100100K
- **2012-10-19:** UCI Legal Case Reports
- **2012-09-29:** UCI seeds
- **2012-08-30:** UCI Individual household electric power consumption

Most Popular Data Sets (hits since 2007):
- **434000:** Iris
- **304178:** Adult
- **264595:** Wine
- **219860:** Breast Cancer Wisconsin (Diagnostic)
- **207070:** Car Evaluation
- **189658:** Abalone
- **148618:** Poker Hand
- **125417:** Forest Fires
- **120653:** Wine Quality
- **110154:** MONK's Problems
- **109832:** Internet Advertisements
Weka 3: Data Mining Software in Java

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Found only on the islands of New Zealand, the Weka is a flightless bird with an inquisitive nature. The name is pronounced like this, and the bird sounds like this.

Weka is open source software issued under the GNU General Public License.

---

Getting started

- Requirements
- Download
- Documentation
- FAQ
- Getting Help

Further Information

- Citing Weka
- Datasets
- Related Projects
- Miscellaneous Code
- Other Literature

Developers

- Development
- History
- Subversion
- Contributors
Challenges

• Variety
  – We need examples from a large suite of domains where preference handling and decision making research happens.

• Elicitation
  – How do we collect this data – how do we ensure the quality of the data we receive?

• Modeling
  – What are the correct formalisms to represent preferences?
Challenges

• Over-fitting
  – How do we prevent becoming too focused on metrics that can be measured from our datasets?

• Privacy and Information Silos
  – Some data cannot be shared, some data is not shared because it may create a competitive advantage – how do we convince people to share?
PrefLib: A Library for Preferences

A reference library of preference data and links assembled by Nicholas Mattei and Toby Walsh. This site and library is proudly supported by the Optimization Research Group at NICTA. We currently house over 3,000 datasets for use by the community.

We want to provide a comprehensive resource for the multiple research communities that deal with preferences, including computational social choice, recommender systems, data mining, machine learning, and combinatorial optimization, to name just a few.

Please see the about page for information about the site, contacting us, and our citation policy. We rely on the support of the community in order to grow the usefulness of this site. To contribute, please contact Nicholas Mattei at: nicholas[dot]mattei@nicta.com.au

Supported By:

NICTA

Links
- UC Irvine Machine Learning Repository
- University of Minnesota GroupLens Data Sets
- CSPLib: A Problem Library for Constraints
- Microsoft Learning to Rank Datasets
- SATLib: The Satisfiability Library
- Preference-Learning.org
- Toshihiro Kamishima’s Sushi Preference Datasets
- MAX-SAT Evaluations and Datasets

Dec. 10, 2013:
We are hosting a workshop at AAMAS 2014 on Exploring Beyond the Worst Case inComputational Social Choice. Nick will give a talk about PrefLib! Please consider joining us in Paris in the coming year.

Nov. 6, 2013:
The first release of the tool suite is now available on the Tools page. Python3 scripts to read, write, and generate preference data in our formats!

Sept. 3, 2013:
A big update today brings us over 3000 datasets hosted on the site with a full data archive over 7 GB!

We have also added a Thanks! section to recognize those individuals who have helped make PrefLib possible.

July 1, 2013:
Our paper has been accepted to 2013 Conference on Algorithmic Decision Theory. We have also had several new donated datasets which have been parsed and posted.
Visitors – A Year On

Sessions: 1,532
Pageviews: 6,597
Users: 894
Pages / Session: 4.31

New Visitor: 57.8%
Returning Visitor: 42.2%
Visits by Country
Types of Data

• We divide our data into 4 broad categories:
  – Election Data
  – Matching Data
  – Combinatorial Data
  – Optimization Data

• Simple data formats, mostly derived from comma separated lists.
  – As SATLib showed, this format is easy to use across software and disciplines.
Election Data

• Elections, rank-ordered preferences, partial-orders, sports competitions, tournaments, majority graphs.

\[ a > b > c > d \]

• Datasets ranging from NASA satellite path selection to international skating competitions and real elections

\[ a > b, c, d > e \]
Sample Data Format

Number of Candidates: 11
Candidate List: 1, Australia, 2, Braille, 3, Brush Strokes, 4, Exponential, 5, College, 6, Graph Coloring, 7, Red, 8, Simple, 9, Star Trek, 10, TSP, 11, VRP
Number of Voters, Number of Orders, Number of Unique Orders: 30, 30, 30
Comma Separated Orders:
- 1, 10, 6, 7, 8, 11, 5, 3, 2, 1, 9, 4
- 1, 1, 10, 11, 9, 6, 7, 3, 5, 8, 2, 4
- 1, 11, 10, 1, 3, 6, 8, 5, 7, 4, 9, 2
- 1, 10, 3, 5, 8, 11, 1, 6, 2, 4, 7, 9
Election Data – Key Stats

• 14 Unique Data Sources
  – Irish Elections, Scottish Elections, Course Selection, Skate Rankings, Movies, Sushi

• 470+ different election instances with strict rankings
  – 229 Contain only complete rankings.
  – 243 Contain some incomplete rankings.

• 381 different election instances where rankings have indifference
Voting Datasets

Number of Voters

Number of Candidates
Voting Datasets

Number of Voters

Number of Candidates
Voting Datasets

Number of Voters vs Number of Candidates

Nicholas Mattei
NICTA and UNSW
Models – Strict (SOC)
Models – Strict (SOI)
Models – Partial (POC)

>  

>  

JAWS

Transformers

The Godfather

Plan 9 from Outer Space
Models – Incomplete (POI)
Not Uncommon…. 

• Irish Election Data:  
  – 5% submitted complete ballots for Dublin North.  
  – 12% for Dublin West. 

• APA Election Data  
  – Chamberlin’s original data had over 65% incomplete ballots over 5 candidates. 

• ANES Thermometer Rankings  
  – Takes ratings and turns them into rankings, breaking ties randomly. 

• Sushi Dataset  
  – Incomplete survey’s are discarded (sample bias, not incentivized).
Problems with Extensions

• Behavioral aspects of individuals can have substantial impacts on the resulting computational problems
  – Youtube’s dropping of the star ratings system…
  – Single-peakedness…

• As we move forward pay particular attention to the domain in which we wish to deploy our results.

• We must be acutely aware of model dependence.
Models - Agnostic

> 

> 

>
Models - Pessimistic
Models – Anchor and Adjust

Nicholas Mattei
NICTA and UNSW
Some Extensions We Provide

- All have been turned into pairwise majority graphs, weighted pairwise majority, and tournament graphs.

- We have not posted all the different types of model extensions possible – we want these extensions to be explicit.

- Our goal is to present the data and preserve the context for future researchers – avoid over-fitting!
Combinatorial Data

- CP-nets, UAI-nets, single and multi-attribute rating data.

- Currently we have a large collection of data scraped from TripAdvisor.
Audience Question

• New Category?
  – Rating / Multi-Dimensional Preference Data?
### Sample Data Format

<table>
<thead>
<tr>
<th>Hotel ID</th>
<th>User ID</th>
<th>Price</th>
<th>Location</th>
<th>Overall Rating</th>
<th>Value</th>
<th>Service Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>100504</td>
<td>selizabethm</td>
<td>302</td>
<td>Seattle Washington</td>
<td>5,4,5,5,5,5,5</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>100504</td>
<td>IndieLady</td>
<td>302</td>
<td>Seattle Washington</td>
<td>4,5,4,5,4,5,5</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>100504</td>
<td>H'obb</td>
<td>302</td>
<td>Seattle Washington</td>
<td>4,4,4,3,4,5,1</td>
<td>-1,4</td>
<td></td>
</tr>
<tr>
<td>100504</td>
<td>Ch. antennigirl24</td>
<td>302</td>
<td>Seattle Washington</td>
<td>5,5,5,5,5,5,5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100504</td>
<td>hothearted</td>
<td>302</td>
<td>Seattle Washington</td>
<td>5,-1,-1,-1,-1,-1,-1</td>
<td>-1,1</td>
<td></td>
</tr>
<tr>
<td>100504</td>
<td>MauiDiver</td>
<td>302</td>
<td>Seattle Washington</td>
<td>2,2,3,3,5,2,2,3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Matching Data

• Kidney donor matching, runway slots, student seats in classrooms, student housing.

• Currently we have synthetic kidney matching data and student preferences over courses.
Sample Data Format

16,59
1,Pair 1
2,Pair 2
3,Pair 3
4,Pair 4
5,Pair 5
...
16,Pair 16

Number of Nodes
Number of Edges

Item Lists

Source

Destination

Weight

0,4,1
0,5,1
1,11,1
1,10,1
1,0,1
1,2,1
2,4,1
2,7,1
Optimization Data

• Max-SAT, Max-CSP, weighted-CSP, TSP-games, etc.

• We lack good data here!!
Optimization Data

- Max-SAT,
  Max-CSP,

Max-SAT 2014

Ninth Max-SAT Evaluation

Introduction

The Ninth Evaluation of Max-SAT Solvers (Max-SAT-2014) is organized as an affiliated event of the 17th International Conference on Theory and Applications of Satisfiability Testing (SAT-2014).

The objective of the evaluation is assessing the state of the art in the field of Max-SAT solvers, as well as creating a collection of publicly available Max-SAT benchmark instances.
**Tools**

- We have released version 0.1 of our tool suite – read, write, and convert between all the election data instances and matching data instances.

- I’ll show some features of 0.2 which will accompany a new data update around 1 June.
Major Tool Components

• PreflibUtils.py: Main library of with functions for reading, writing, printing, and manipulating the data files.

• GenProfiles.py: Main file for generating preferences. Thanks to Andreas Pfandler for help with updates!

• DomainRestrictions.py: Test for various Domain restrictions including Single Peakedness.
Statistical Cultures

- The **Impartial Culture** assumes that the probability of observing any order is uniform.

<table>
<thead>
<tr>
<th>A &gt; B &gt; C</th>
<th>A &gt; C &gt; B</th>
<th>B &gt; A &gt; C</th>
<th>B &gt; C &gt; A</th>
<th>C &gt; A &gt; B</th>
<th>C &gt; B &gt; A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
</tr>
</tbody>
</table>
Statistical Cultures

• The **Impartial Anonymous Culture** assumes that the probability of observing any distribution over orders is equally likely.

• That is, any vector where
  \[ X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 1.0 \]
Statistical Cultures

• The **Urn Model** starts with a bag with $N!$ elements in it.

<table>
<thead>
<tr>
<th>A &gt; B &gt; C</th>
<th>A &gt; C &gt; B</th>
<th>B &gt; A &gt; C</th>
<th>B &gt; C &gt; A</th>
<th>C &gt; A &gt; B</th>
<th>C &gt; B &gt; A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
</tr>
</tbody>
</table>

• We draw a vote from it, and replace it with some fixed number of copies of the vote.
Statistical Cultures

• The **Urn Model** starts with a bag with $N!$ elements in it.

<table>
<thead>
<tr>
<th>A &gt; B &gt; C</th>
<th>A &gt; C &gt; B</th>
<th>B &gt; A &gt; C</th>
<th>B &gt; C &gt; A</th>
<th>C &gt; A &gt; B</th>
<th>C &gt; B &gt; A</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
</tr>
</tbody>
</table>

• We draw a vote from it, and replace it with some fixed number of copies of the vote.
  – Replace with 1 extra copy.
Statistical Cultures

• **Mallows Models** are parameterized by a reference ranking and a dispersion parameter.

• The probability of observing the non-reference ranking is proportional to the normalized value of the dispersion parameter ($\phi$) raised to the swap distance.

<table>
<thead>
<tr>
<th>A &gt; B &gt; C</th>
<th>A &gt; C &gt; B</th>
<th>B &gt; A &gt; C</th>
<th>B &gt; C &gt; A</th>
<th>C &gt; A &gt; B</th>
<th>C &gt; B &gt; A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Statistical Cultures

- \( P(C > B > A) = \frac{1}{X} \phi^{KT(A>B>C,C>B>A)} \)
- With \( X = 1 \cdot (1 + \phi) \cdot (1 + \phi + \phi^2) \ldots (1 + \ldots + \phi^{m-1}) \)

- So this means that as \( \phi \to 1 \) we get IC
- And as \( \phi \to 0 \) we get a unanimous distribution.

- So \( X = 2.625, \phi = 0.5 \)

<table>
<thead>
<tr>
<th>( A &gt; B &gt; C )</th>
<th>( A &gt; C &gt; B )</th>
<th>( B &gt; A &gt; C )</th>
<th>( B &gt; C &gt; A )</th>
<th>( C &gt; A &gt; B )</th>
<th>( C &gt; B &gt; A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.381</td>
<td>0.1905</td>
<td>0.1905</td>
<td>0.095</td>
<td>0.095</td>
<td>0.04762</td>
</tr>
</tbody>
</table>
Statistical Cultures

- **Single Peaked Impartial Culture** draws uniformly at random from the set of orders that are consistent with a given social axis.

- If the axis is $A > B > C$ then we have positive probability only on the $2^{m-1} = 2^2 = 4$ consistent orders.

<table>
<thead>
<tr>
<th>$A &gt; B &gt; C$</th>
<th>$A &gt; C &gt; B$</th>
<th>$B &gt; A &gt; C$</th>
<th>$B &gt; C &gt; A$</th>
<th>$C &gt; A &gt; B$</th>
<th>$C &gt; B &gt; A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/4</td>
<td>0</td>
<td>1/4</td>
<td>1/4</td>
<td>0</td>
<td>1/4</td>
</tr>
</tbody>
</table>
Matplotlib and PyLab

• Available as a combined distribution with Numpy and other libraries at http://scipy.org/
• Easy to use and intuitive way to make graphs and do end to end experiments.
• Python is magic…

SciPy.org

Install Getting Started Documentation Report Bugs
Demo Time!
New Data Experiments

• Using the IC model and IC generalized to partial rankings we asked how much PrefLib looks like generative models.
  – Should we define a discrete probability distribution called the PrefLib Culture?

• For comparison sake we randomly create additional datasets:
  – 50% Off: Each observed probability is off from IC by 50%. For 3 candidates: \(P(A > B > C) = 0.25\) instead of 0.16667.

  – 50% Missing: There is probability 0 of predicting 50% of the observations. \(P(A > B > C) = 0, P(C > B > A) = 0.333\).
Comparing IC – Complete Orders

- **Average Euclidian Error**
- **Number of Candidates**

Legend:
- Avg. Error
- Off 50%
- 50% Missing
Comparing IC – Incomplete Orders

- Avg. Error
- Off 50%
- 50% Missing

Nicholas Mattei
NICTA and UNSW
IC – Incomplete Orders – Log Scale

Graph showing the relationship between Avg. Euclidian Error and Number of Candidates, with different error rates represented.

- Avg. Error
- Off 50%
- 50% Missing
All Sets – Error by Candidate Count

Average Euclidian Error

Number of Candidates

Nicholas Mattei
NICTA and UNSW
What to do Now?

• Now that we have more data, what can we do in decision theory that we couldn’t before?

  – Learning the preferences of one or multiple users in ordinal environments?

  – Learning and modeling preference drift over time.

  – Extending preferences from incomplete to complete orders.
Thanks

- Everyone who has participated and donated data or their time to the project so far:
  - Robert Bredereck
  - John Dickerson
  - Carleton Coffrin
  - Piotr Faliszewski
  - Toshihiro Kamishima
  - Jeffrey O'Neill
  - Andreas Pfandler
  - Florenz Plassmann
  - Nicolaus Tideman
  - Hongning Wang
Questions for You!

• Continue to grown the community and types of data. Please tell us if you have data to share!!

• Other key tools or pieces that we should consider adding?

• Any issues with the data-formats we are proposing?

• What kind of data would this community like to see?
Thanks!

• Questions

• Comments