Content-Based Recommendation

Adaptive Web Systems
Peter Brusilovsky
With slides from Dr. Jae-wook Ahn
Content-based?

• Item *descriptions* to identify items that are of particular interest to the user
Example
American boxers will go home from an Olympics without a gold medal for the first time since 1948. Ricardo Williams Jr. of Cincinnati, who had outpointed Diogenes Luna of Cuba 42-41 in a semifinal bout, settled for a silver medal when he lost 27-20 to Abdullaev Mahamadkadyz of Uzbekistan in the 139-pound final. World champion Rocky Juarez of Houston could not get inside enough against a clutching, grabbing Bekzal Sattarkanov of Kazakhstan, and lost 22-14 at 125 pounds. The defeat snapped the 20-year-old Juarez' two-year winning streak at 68 bouts. ``I did all I could do, but it wasn't good enough,'' Juarez said. ``I didn't come here to get the silver medal. I'm disappointed.''

Juarez complained about Sattarkanov's holding tactics throughout the bout. ``I think he should have been disqualified,'' Juarez said.

U.S. team manager Gary Toney said he would file a protest, charging that Russian referee Stanislav Kirsanov cautioned Sattarkanov nine times but never issued a warning that could have cost the fighter points or led to a disqualification. ``I have no idea why the referee was allowing it,'' said Toney, who acknowledged the appeal would likely fall through. Two Americans earned bronze medals - Clarence Vinson of Washington, D.C., at 119 pounds and Jermain Taylor of Little Rock, Ark., at 156 pounds. The four medals are two less than Americans won in 1996 at Atlanta (one gold and five bronze) and one more than they collected in 1992 at Barcelona (one gold, one silver, one bronze).

Cuba, which had no boxers in Sunday's six finals, matched its Atlanta total of four golds on Saturday. One of them was the third for heavyweight Felix Savon. Two Cubans also got two bronze medals. The 5-foot-3 Juarez, four inches shorter than his opponent, got hit repeatedly by left hands and trailed 15-4 after two rounds. Then trailing 17-8 in the second round, Juarez landed five of the next six scoring blows, but Sattarkanov got home two scoring punches in the closing seconds for a 20-13 lead. Juarez kept charging forward and Sattarkanov kept wrapping him up in the final round.
Comparing with Non-content based

- User-based CF
  Searches for similar users in user-item "rating" matrix

- No rating

- Item-feature matrix
Item Representation

- Structured
- Unstructured
- Semi-structured
Structured

- Attribute-value

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Cuisine</th>
<th>Service</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>Mike’s Pizza</td>
<td>Italian</td>
<td>Counter</td>
<td>Low</td>
</tr>
<tr>
<td>0002</td>
<td>Chris’s Cafe</td>
<td>French</td>
<td>Table</td>
<td>Medium</td>
</tr>
<tr>
<td>0003</td>
<td>Jacques Bistro</td>
<td>French</td>
<td>Table</td>
<td>High</td>
</tr>
</tbody>
</table>
Unstructured

• Full-text
• No attributes formally defined
• Other complicated problems - such as synonymy, polysymy
Semi-structured

- Structured + unstructured
- Well defined attributes/values + free text
Conversion of Unstructured Data

- Need to convert to structured form
- IR techniques
- VSM, TF-IDF, stemming, etc.
User Profile

• A model of user’s preference
• A history of the user’s interactions
  • Recently viewed items
  • Already viewed items
• Training data — machine learning
User Profiling

Adaptive System

- Collects information about individual user
- Provides adaptation effect

User Modeling side

User Model

Adaptation side

User profiling == user modeling
User Profile

- Customization
- Manual
- Limitations — user efforts (interest change), no ordering
- Rule-based recommendation — history based rules
User Profile

Edit Favorites

Mark the categories that interest you the most.

- [ ] Books
- [x] Biographies & Memoirs
- [x] Business & Investing
- [x] Computers & Internet
- [ ] Nonfiction

Add to Your Favorites

- [ ] Arts & Photography
- [ ] Craft & Hobbies
- [ ] Romance
- [ ] Comics & Graphic Novels
- [ ] Cooking, Food & Wine
- [ ] Travel
- [ ] Outdoors & Nature
- [ ] Parenting & Relationships
- [ ] Professional & Technical
- [ ] Reference
- [ ] Science & Math
- [ ] Sports & Outdoors
- [ ] Travel

Submit
User Model Learning

• Classification problem
• Classifying to “Like” or “Dislike”
• Training data — feedbacks
• Probability of classification
• Unstructured data conversion
• Feature selection — high/low dimensions
User Model Learning

Training Data
- Items (Like)
- Items (Dislike)

Target Data
- Items
- Items (Like)
- Items (Dislike)

Train/Learn

Classify/Recommend

User Model
UM Learning Feedbacks

- Implicit feedback
- Indirect interaction
- Opened document, Reading time, etc.
- Large data, high uncertainty
Explicit

Directly from users

No noise, hard to obtain
User Model Learning
Feature Selection

• Problem of high dimensional input vectors

• Overfit (especially when a dataset is small)

• Document frequency thresholding, Information gain, Mutual information, Chi square statistic, Term strength
Overfitting

- Overfit
- Underfit

![Graph showing Overfit and Underfit](image-url)
• **Mutual Information**

- \( A = \) number of times \( t \) and \( c \) co-occur
- \( B = \) number of times \( t \) occurs without \( c \)
- \( C = \) number of times \( c \) occurs without \( t \)
- \( N = \) number of total documents

\[
I(t, c) = \log \frac{A \times N}{(A + C) \times (A + B)}
\]
User Model Learning
Feature Selection

- “Austrian train fire accident”
- After learning 5 documents

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Fire</th>
<th>Alps</th>
<th>Austria</th>
<th>People</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>5873</td>
<td>8092</td>
<td>93</td>
<td>974</td>
<td>34501</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>96260</td>
<td>96260</td>
<td>96260</td>
<td>96260</td>
<td>96260</td>
</tr>
</tbody>
</table>
Decision Tree

- Partitioning dataset into trees
- Ideal for **structured, small** data
- Performance, simplicity, understandability
- ID3 (Iterative Dichotomiser 3)
• Using WEKA
• http://www.cs.waikato.ac.nz/ml/weka/
• Machine learning algorithm package
• JAVA API, interactive UI
**Decision Tree**

**Example - dataset**

**Training**

```
@relation lensesTrain

@attribute Age {young, pre-presbyopic, presbyopic}
@attribute Prescription {myope, hypermetrope}
@attribute Astigmatic {no, yes}
@attribute Tear_rate {normal, reduced}
@attribute Lenses {YES, NO}

@data

young, myope, no, normal, YES
young, myope, no, reduced, NO
pre-presbyopic, myope, no, normal, YES
pre-presbyopic, hypermetrope, no, reduced, NO
pre-presbyopic, myope, yes, normal, YES
pre-presbyopic, myope, yes, reduced, NO
```

**Testing**

```
@relation lensesTest

@attribute Age {young, pre-presbyopic, presbyopic}
@attribute Prescription {myope, hypermetrope}
@attribute Astigmatic {no, yes}
@attribute Tear_rate {normal, reduced}
@attribute Lenses {YES, NO}

@data

young, hypermetrope, no, normal, YES
pre-presbyopic, myope, no, normal, YES
pre-presbyopic, hypermetrope, no, reduced, NO
```

**attribute** as inputs

**attribute** to be estimated
Decision Tree

Example - tree

Classifier

Choose Id3

Test options

- Use training set
- Supplied test set
- Cross-validation
- Percentage split

More options...

Classifiers output

=== Classifier model (full training set) ===

Id3

Tear_rate = normal
  Age = young: YES
  Age = pre-presbyopic: YES
  Age = presbyopic
    Prescription = myope
      Astigmatic = no: NO
      Astigmatic = yes: YES
    Prescription = hypermetrope: YES
Tear_rate = reduced: NO

Time taken to build model: 0 seconds

Predictions on test set ===

<table>
<thead>
<tr>
<th>Inst#,</th>
<th>actual, predicted, error, probability distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1:YES 1:YES</td>
</tr>
<tr>
<td>2</td>
<td>1:YES 1:YES</td>
</tr>
<tr>
<td>3</td>
<td>2:NO 2:NO</td>
</tr>
<tr>
<td>4</td>
<td>1:YES 1:YES</td>
</tr>
<tr>
<td>5</td>
<td>2:NO 1:YES</td>
</tr>
<tr>
<td>6</td>
<td>2:NO 2:NO</td>
</tr>
<tr>
<td>7</td>
<td>2:NO 1:YES</td>
</tr>
</tbody>
</table>
Decision Tree Example - tree

Tear Rate
- Reduced
- Normal

Age
- Young
- Pre-presbyoptic

Prescription
- Yes
- Presbyoptic
- Hypermetrope
- Myope

Astigmatic
- Yes
- No
### Evaluation on test set

#### Summary

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly Classified Instances</td>
<td>5</td>
<td>71.4286 %</td>
</tr>
<tr>
<td>Correctly Classified Instances</td>
<td>2</td>
<td>28.5714 %</td>
</tr>
<tr>
<td>$\kappa$ statistic</td>
<td>0.4615</td>
<td></td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>0.2857</td>
<td></td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>0.5345</td>
<td></td>
</tr>
<tr>
<td>Local relative squared error</td>
<td>59.375  %</td>
<td></td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

#### Detailed Accuracy By Class

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.6</td>
<td>1</td>
<td>0.75</td>
<td>0.75</td>
<td>YES</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>0.5</td>
<td>0.667</td>
<td>0.75</td>
<td>NO</td>
</tr>
</tbody>
</table>

#### Weighted Avg.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted Avg.</td>
<td>0.714</td>
<td>0.214</td>
<td>0.829</td>
<td>0.714</td>
<td>0.702</td>
<td>0.75</td>
</tr>
</tbody>
</table>

### Confusion Matrix

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>a = YES</td>
</tr>
<tr>
<td>1</td>
<td>b = NO</td>
</tr>
</tbody>
</table>
k Nearest Neighbor

• Prepare training data (classification labels)
• Extract k most similar items
• Decide the label of the test data by looking at kNN’s
k Nearest Neighbor

Example

k=N Classified as \textcolor{red}{red}

k=5 Classified as \textcolor{blue}{blue}
Linear Classifier

- Tries to find out a hyperplane that best separates classes
- Gradient Descent - incremental vector
- Support Vector Machine
- Maximizes the distance between decision boundary & support vector (closest training instance)
- Avoids overfitting
Naive Bayes

- VSM — lack of theoretical justification
- Probabilistic text classification method
- Naive = term independence
- Probability document $d$ is classified to category $c$

$$P(c_j | d; \hat{\theta}) = \frac{P(c_j | \hat{\theta}) P(d_i | c_j; \hat{\theta})}{\sum_{j} P(c_j | \hat{\theta}) P(d_i | c_j; \hat{\theta})}$$
Naive Bayes

- Multivariate Bernoulli
- Document probability = $\prod$ of term probability
  $\rightarrow$ term independence assumption
  $\rightarrow$ “naive”

\[
P(d_i \mid c_j; \theta) = \prod_{t=1}^{V} (B_{it} P(w_t \mid c_j; \theta) + (1 - B_{it})(1 - P(w_t \mid c_j; \theta)))
\]
Naive Bayes

- Multinomial
- Non-binary

\[
P(d_i \mid c_j; \theta) = P(|d_i|) \prod_{t=1}^{d_i} P(w_t \mid c_j; \theta)^{N_{it}}
\]

\[
P(w_t \mid c_j; \theta) = \frac{1 + \sum_{i=1}^{D} N_{it} P(c_j \mid d_i)}{|V| + \sum_{i=1}^{D} \sum_{s=1}^{V} N_{is} P(c_j \mid d_i)}
\]
### Naive Bayes Example

#### Classifier output

```markdown
--- Predictions on test set ---

<table>
<thead>
<tr>
<th>inst#, actual, predicted, error, probability distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1           1=YES 1=YES 0.833 0.167</td>
</tr>
<tr>
<td>2           1=YES 1=YES 0.722 0.278</td>
</tr>
<tr>
<td>3           2:NO 2:NO 0.097 0.903</td>
</tr>
<tr>
<td>4           1=YES 1=YES 0.752 0.248</td>
</tr>
<tr>
<td>5           2:NO 1=YES + 0.829 0.171</td>
</tr>
<tr>
<td>6           2:NO 2:NO + 0.112 0.888</td>
</tr>
<tr>
<td>7           2:NO 1=YES + 0.784 0.216</td>
</tr>
</tbody>
</table>
```
Conclusions

• User model learns from content (description, fulltext, etc) itself

• Implicit, explicit method

• Classifying — like, dislike

• Limitations — not enough content

• Hybrid — content + collaborative