Workshop: Automated Competence Assessment
Frejus, May 31, 2007

Fridolin Wild
Vienna University of Economics and Business Administration

Workshop Plan
- Introduction: Competence, Autom. Assessment
- Example 1: Mining Social Competence with Latent Semantic Analysis
- Example 2: Essay Scoring with Latent Semantic Analysis
- Example 3: Positioning & Accreditation of Prior Knowledge (Group Formation Example)
- Example 4: Analysing Professional Communities with Social Network Analysis (Tag-Person Nets)
- Open Space

Introduction

The History of Competence

... just a selection ...

Competence Definition

“A competence is defined as the ability to successfully meet complex demands in a particular context through the mobilization of psychosocial prerequisites (including both cognitive and noncognitive aspects)”
(Rychen & Salganik, 2003b, p. 43)

Competence Dimensions

- Competence is a human potentiality for action
- … is demand oriented (= abilities required for e.g. task)
- … refers to abilities that can be learned
- … involves cognitive and non-cognitive elements:
  - factual knowledge
  - procedural skills
  - internalised orientations
  - values
  - attitudes
  - volitional aspects
  - …
Competence Classes (I)

- Excerpted from empirical, political, and theoretical perspectives (see paper) ...
- **Professional competence**
  - basic and specialized general knowledge, basic psychomotor and mechanical skills, and disciplinary and interdisciplinary knowledge (Jäger, 2001)
- **Methodological competence**
  - ability to independently acquire, structure, critically evaluate, and exploit knowledge in a creative way (Kauffeld et al., 2003)

Competence Classes (II)

- **Social Competence**
  - facilitate communicative and cooperative action and that aim at identifying, managing and mastering conflicts (Erpenbeck, 2003)
- **Personal Competence**
  - concerned with those attitudes and character attributes required to perceive and utilize one’s own competencies and to act in a reflective and self-reflective way (Erpenbeck, 2003)

Important Competences

Automated Measurement

Automated Measurement (II)

- **Natural Language Processing Approaches (NLP)**
  - Syntax-based: structural analysis regardless meaning
    - Shallow counting (orthography, e.g. Page, 1966)
    - Structural Analysis (e.g. POS-tagger & discourse structure parser)
  - Semantics-based: analysis of the meaning
    - Concept-based
    - Context-based
Social Network Analysis

### Existing for a long time (term coined 1954)

### Basic idea:
- Actors and Relationships between them (e.g. Interactions)
- Actors can be people (groups, media, tags, ...)
- Actors and Ties form a Graph (edges and nodes)
- Within that graph, certain structures can be investigated
  - Betweenness, Degree of Centrality, Density, Cohesion
  - Structural Patterns can be identified (e.g. the Troll)

### Input Data: Interactions

<table>
<thead>
<tr>
<th>authorparent_id</th>
<th>forum_id</th>
<th>message_id</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Adjacency Matrix

<table>
<thead>
<tr>
<th></th>
<th>Frido</th>
<th>Stefan</th>
<th>Gustaf</th>
<th>Steinn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frido</td>
<td>2</td>
<td>3</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Stefan</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Gustaf</td>
<td>-</td>
<td>4</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Steinn</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### Analysis of a Discussion Board

- Message Board: Business English
- Most central Author 1083 (Highest Degree Centrality, Highest Betweenness)
- => a student!

### Social Proxies

**Babble**

(Staples, 1999)
Personal Networks & Community Networks

Analysis of Co-Authorships in Prolearn:
1st Order Personal Network

Analysis of Co-Authorships in Prolearn:
2nd Order Personal Network

Analysis of Co-Authorships in Prolearn:
The Structure of the Community

Tag-Person Networks
(Cumulative Wisdom)

Tag-Person Network (iCamp)
Latent Semantic Analysis

- Assumption: documents have a semantic structure
- Structure is obscured by word usage (noise, synonyms, homographs, …)
- Therefore: map text matrix using conceptual indices derived statistically (truncated SVD):

\[ M_2 = (T)(S_2)(D)' \]

Input (e.g., documents)

- Only the red terms appear in more than one document, so strip the rest.

Singular Value Decomposition

\[ M = T S D^T \]

Latent Semantics

\[ T_k S_k D_k^T = M_k \]
Similarity in a Latent-Semantic Space

Ex Post Updating: Folding-In
- SVD factor stability
  - SVD calculates factors over a given text base
  - Different texts – different factors
  - Challenge: avoid unwanted factor changes (e.g., bad essays)
  - Solution: folding-in of essays instead of recalculating
- SVD is computationally expensive
  - 14 seconds (300 docs textbase, this machine)
  - 10 minutes (3500 docs textbase, this machine)
  - ... and rising!

Analogy to Humans
- “Humans learn word meanings and how to combine them into passage meaning through experience with ~paragraph unitized verbal environments.”
- “They don’t remember all the separate words of a passage; they remember its overall gist or meaning.”
- “LSA learns by ‘reading’ ~paragraph unitized texts that represent the environment.”
- “It doesn’t remember all the separate words of a text it; it remembers its overall gist or meaning.”

Word Choice
- Educated adult understands ~100,000 word forms
- An average sentence contains 20 tokens.
- Thus 100,000\(^2\) possible combinations of words in a sentence
- Therefore: maximum of \(\log_2 100,000^2\) = 332 bits in word choice alone.
- \(20! = 2.4 \times 10^{18}\) possible orders of 20 words = maximum of 61 bits from order of the words.
- \(332/(61+332) = 84\% \) word choice

The meaning of “life” =

LSA Process & Driving Parameters
Working Principle

(Dumb) Essay Scoring

Evaluating Effectiveness

- Compare Machine Scores with Human Scores
- Human-to-Human Correlation
  - Usually around .6
  - Increased by familiarity between assessors, tighter assessment schemes,
  - Scores vary even stronger with decreasing subject familiarity (.8 at high familiarity, worst test -.07)

Essay Scoring (Code)

```r
library( "lsa" )

# load package
# load training texts
trm = textmatrix( "trainingtexts/" )
trm = lw_bintf( trm ) * gw_idf( trm )
# weighting
space = lsa( trm )
# create an LSA space

# fold-in essays to be tested (including gold standard text)
tem = textmatrix( "testessays/", vocabulary=rownames(trm) )
tem_red = fold_in( tem, space )
# score an essay by comparing with
# gold standard text (very simple method!)
cor( tem_red[,"goldstandard.txt"], tem_red[,"E1.txt"], )
=> 0.7
```

(Positive) Evaluation Results

LSA machine scores:

<table>
<thead>
<tr>
<th></th>
<th>Spearman's rank correlation rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>hyp</td>
<td>alternative hypothesis: true rho is not equal to 0</td>
</tr>
<tr>
<td>sample estimates:</td>
<td>rho</td>
</tr>
<tr>
<td></td>
<td>0.687324</td>
</tr>
</tbody>
</table>

Pure vector space model:

<table>
<thead>
<tr>
<th></th>
<th>Spearman's rank correlation rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>hyp</td>
<td>alternative hypothesis: true rho is not equal to 0</td>
</tr>
<tr>
<td>sample estimates:</td>
<td>rho</td>
</tr>
<tr>
<td></td>
<td>0.4447588</td>
</tr>
</tbody>
</table>

Demo (User)
Mining Social Competence

Selection (!) of Dimensions of Social Competence
- Empathy, Politeness
- Phatic communication
- Ability to express own opinion
- Cooperation competence
- Team competence
- Ability to take initiative
- Ability to motivate
- Readiness to take on responsibility

Our Approach
- Use forum messages that have been evaluated by humans on whether they contain certain dimensions of social competence to classify new forum messages by assigning dimensions whenever the weighted sum of the positive occurrences is higher than that of the absent best hits

Example: Ability to motivate
- Motivate yourself and others (!)
- Examples:
  - „grossartig, dass du es noch geschafft hast!“ ~ great that you still made it!
  - „sonst freu mich eure posts zu lesen; ihr schafft es sicher!“ ~ furthermore, I am happy to read your posts; you will make it for sure!
  - „ich glaube ich schaff das schon.“ ~ I think i will manage to do it.

Research Design (1)
- 337 German contributions from students in one university seminar forum
- Split into 1,012 sentences = corpus
- Coded by human assessors along ten dimensions of social competence (one dim dropped out with only 37 messages)
- Corpus was split into 490 training and 522 test texts
- Of the 522, 16 were omitted (no terms from the training space)
Occurrences (Human, All)
- Politeness: 197
- Phatic communication: 579
- Networking Competence: 113
- Ability to express own opinion: 156
- Cooperation competence: 774
- Team competence: 144
- Ability to take initiative: 207
- Ability to motivate others: 074
- Readiness to take on responsibility: 215
- Feedback competence: 362

Research Design (2)
- Space over 490 training docs calculated (dimcalc share 0.5, no stemming & no stopping, minDocFreq = 1, minWL = 0)
- Each of the 506 folded into the space
- Pearson's r to compare against all training documents
- For each doc, 10 highest correlating docs were selected
- Sum up correlations of 'positive' and 'negative' docs
- If cor sum of positives is higher, it will be assumed that the doc indicates this dimension

Singular Values of the Corpus
![Singular Values Graph]

Example: Politeness of T011.txt
- polite = 1 for the documents
  - D015.txt (r = 0.7597227)
  - D133.txt (r = 0.7597227)
  - D230.txt (r = 0.7597227)
  - D063.txt (r = 0.5035313)
- sum of r's = 2.7826994
- polite = 0 for the documents
  - D311.txt (r = 0.6581647)
  - D126.txt (r = 0.6147096)
  - D055.txt (r = 0.6057419)
  - D034.txt (r = 0.5638606)
  - D299.txt (r = 0.5589148)
  - D031.txt (r = 0.5246403)
- sum of r's = 3.5260319

Results (Percentage of Correct Classifications)

Example: T011.txt is inferred to be polite = 0

Topic-Based Group Formation
Group Formation

- Standard task in the trials
- Important e-tivity
- Several options, how to structure:
  - Based-on similarities (e.g. ProLearn Summer School)
  - Based-on specialisation (mixed teams)
- Usually: assignment by hand
- Alternative: assignment with LSA

Topic-Based Group Formation

- Input documents:
  - Self-Description (CV, Abstract, Paper)
  - Cluster Descriptions
- Take average of all self-descriptions
- Take best-matching cluster as recommendation

Evaluation

- Comparison with Assignment by Hand
- Corpus: ProLearn Summer School Assignments 2006
- Survey among Participants
- Results:
  - Ralf Klamma:
    - 17 matches
    - 12 mismatches
    - 4 didn’t care
    - 6 didn’t answer
  - LSA:
    - 14 matches
    - 15 mismatches
    - 4 didn’t care
    - 6 didn’t answer