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 **psycho-social 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)

<7>

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)

<8>

Important Competences

	Professional Competence	Methodological Competence	Social Competence	Personal Competence
(Gryk, 2003)		Analytical thinking	Communication	Work under pressure
(Schaper & Briedis, 2004)	Domain-specific professional competence	Presentation competence; methodological competence; (self-) organization	Social competence	
(González & Wagenaar, 2003)	Basic general knowledge	Analytic and synthetic capacity to learn; problem-solving		
(European Commission, 2004b)	Mother tongue; Maths	Mother tongue; foreign languages; ICT; numeracy; science and technology; entrepreneurship; learning to learn	Interpersonal and civic competences; general culture	
(European Commission, 2004a)		Problem-solving	Communication; working with others	
(National Postsecondary Education Cooperative, 2004)		Information technology; critical thinking; problem-solving	Communication	
(Weinert, 2001)	Mother tongue; mathematical knowledge	Mother tongue; reading; foreign languages; media; learning strategies; divergent thinking, critical judgment, and self-criticism	Social competences	
(Castle Sperber & Dupuy, 2004)		Perceptive competences; narrative competences	Cooperative competences	Coping with complexity; normative competences
(Rychen & Salganik, 2003a)		Using tools interactively	Joining and functioning in socially heterogeneous groups	Acting autonomously and reflectively

<9>

Automated Measurement

<10>

Automated Measurement

- Four Different Types of Approaches
  - Multiple-Choice Approaches
  - Simulations
    - Virtual labs, online experiments, games
    - From simple click-thru to sophisticated MM
    - Underlying model used to evaluate performance
  - Graph-Based Approaches
    - Based on formalisms such as: concept maps, knowledge maps, mind maps, topic maps, ontologies, Petri nets, adjacency networks, and affiliation networks (plus many others)
    - Mining approaches (e.g. SNA on eMail interaction)
    - Construction approaches (fill-in-the-map vs. construct-a-map)

<11>

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

<12>

Social Network Analysis

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

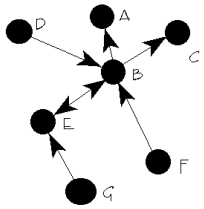
	message_id	forum_id	parent_id	author
130	2853483	2853445	\N	2043
131	1440740	785876	\N	1669
132	2515257	2515256	\N	5814
133	4704949	4699874	\N	5810
134	2597170	2558273	\N	2054
135	2316951	2230821	\N	5095
136	3407573	3407568	\N	36
137	2277393	2277387	\N	359
138	3394136	3382201	\N	1050
139	4603931	4167338	\N	453
140	6234819	6189254	6231352	5400
141	806699	785877	804668	2177
142	4430290	3371246	3380313	48
143	3395686	3391024	3391129	35
144	6270213	6024351	6265378	5780
145	2496015	2491522	2491536	2774
146	4707562	4699873	4707502	5810
147	2574199	2440094	2443801	5801
148	4501993	4424215	4491650	5232

	message_id	forum_id	parent_id	author
60	734569	31117	\N	2491
221	762702	31117		1
317	762717	31117	762702	1927
1528	819660	31117	793408	1197
1950	840406	31117	839998	1348
1047	841810	31117	767386	1879
2239	862709	31117	\N	1982
2420	869839	31117	862709	2038
2694	884824	31117	\N	5439
2503	896399	31117	862709	1982
2846	901691	31117	895022	992
3321	951376	31117	\N	5174
3384	952895	31117	951376	1597
1186	955595	31117	767386	5724
3604	958065	31117	\N	716
2551	960734	31117	862709	1939
4072	975816	31117	\N	584
2574	986038	31117	862709	2043
2590	987842	31117	862709	1982

Adjacency Matrix

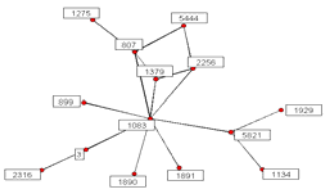
	Frido	Stefan	Gustaf	Steinn
Frido	--	7	4	5
Stefan	5	--	6	4
Gustaf	3	6	--	1
Steinn	5	4	2	--

graph (displayed in a sociogramme)



Analysis of a Discussion Board

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



n	k	no groups
6	2	0
5	2	2
4	2	64
3	2	2691

Calc'ed with k-plex:
- n: number of members to be connected with
- k: number of members no connection is neccessary

Social Proxies

BABBLE

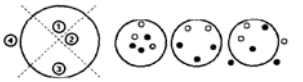


Figure 2. A schematic of the social proxy and, to its right, three instances of it. The schematic shows two people (dots 1 & 2) actively involved in a conversation, one inactive person (dot 3), and one person involved in a different conversation (dot 4). Each dot occupies a (virtual) wedge; wedges are created and destroyed as people log on and off. The first instance shows a 'hot' conversation; the second, a dormant one; the third a mixture of activity, idleness, with three people in other conversations.

(Erickson, 1999)

Personal Networks & Community Networks

<19>

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

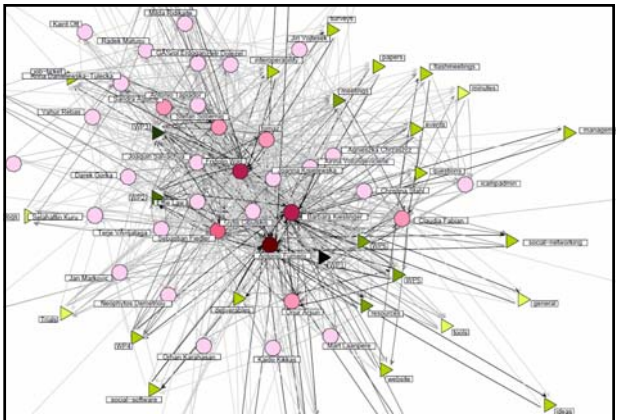
Analysis of Co-Authorships in Prolearn: 2nd Order Personal Network

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

Tag-Person Networks

(Cumulative Wisdom)

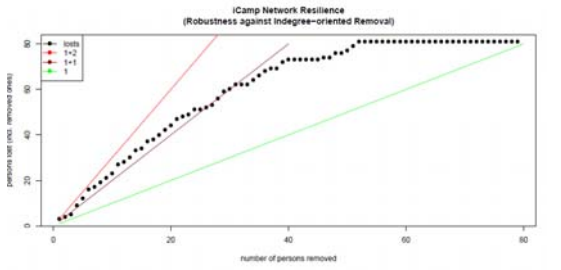
<23>



Tag-Person Network (iCamp)

<24>

Resilience



Latent Semantic Analysis

Latent Semantic Analysis

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

$$\{M_2\} = \{T\}\{S_2\}\{D\}'$$

Input (e.g., documents)

```
c1: Human machine interface for ABC computer applications
c2: A survey of user opinion of computer system response time
c3: The EPS user interface management system
c4: System and human system engineering testing of EPS
c5: Relation of user perceived response time to error measurement

m1: The generation of random, binary, ordered trees
m2: The intersection graph of paths in trees
m3: Graph minors IV: Widths of trees and well-quasi-ordering
m4: Graph minors: A survey
```

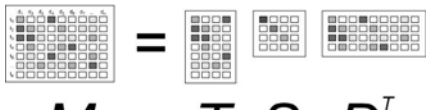
Only the red terms appear in more than one document, so strip the rest.

{ M } =

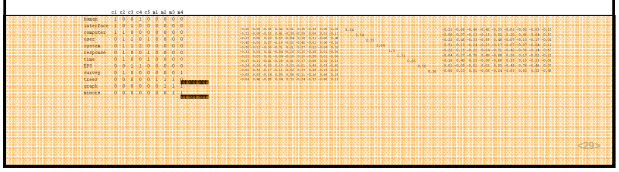
	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

Deerwester, Dumais, Furnas, Landauer, and Harshman (1990): Indexing by Latent Semantic Analysis, In: Journal of the American Society for Information Science, 41(6):391-407

Singular Value Decomposition



$$M = T\ S\ D^T$$

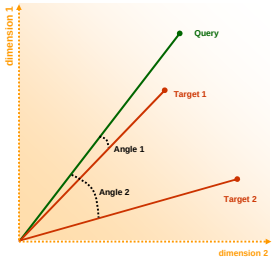


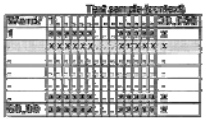
Latent Semantics

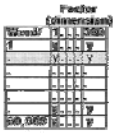
latent-semantic space

$$T_k\ S_k\ D_k^T = M_k$$

Similarity in a Latent-Semantic Space







$$\cos\Theta = \frac{\sum_{i=1}^{m} a_i b_i}{\sqrt{\sum_{i=1}^{m} a_i^2}\sqrt{\sum_{i=1}^{m} b_i^2}}$$

(Landauer, 2007)

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**."

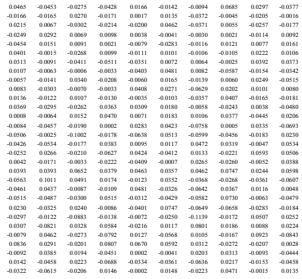
(Landauer, 2007)

Word Choice

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

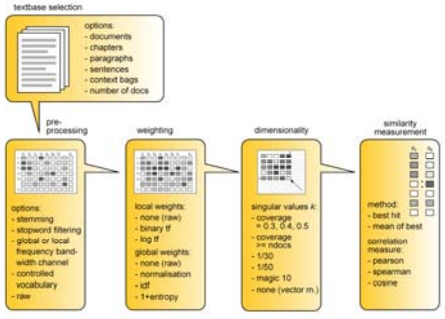
(Landauer, 2007)

The meaning of "life" =



(Landauer, 2007)

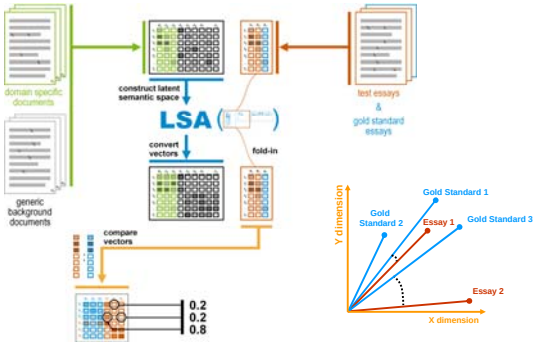
LSA Process & Driving Parameters



Working Principle

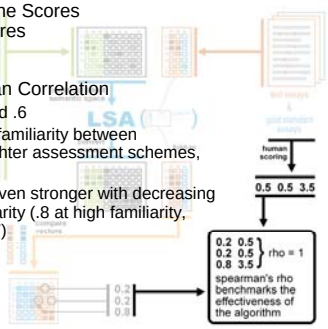
(Landauer, 2007)

‚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)



•Test Collection: 43 German Essays, scored from 0 to 5 points (ratio scaled), average length: 56.4 words
•Training Collection: 3 'golden essays', plus 302 documents from a marketing glossary, average length: 56.1 words

Essay Scoring (Code)

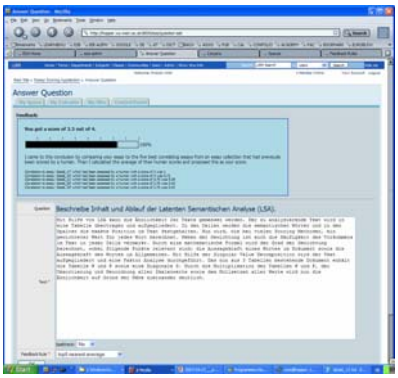
```
library( "lsa" )                          # load package

# 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:
Spearman's rank correlation rho
data: humanscores[names(machinescores),] and machinescores
S = 914.5772, p-value = **0.0001049**
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.687324

Pure vector space model:
Spearman's rank correlation rho
data: humanscores[names(machinescores),] and machinescores
S = 1616.007, p-value = **0.02188**
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.4475188

Demo (User)



Mining Social Competence

Dimensions of Social Competence

(redrawn & translated from Jäger, 2001)

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

(Stahl & Wild, 2006; Ben-Zie, 2004; Langmaack, 2004; Schröder, 1999; Jäger, 2001; Brommer, 1993;

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

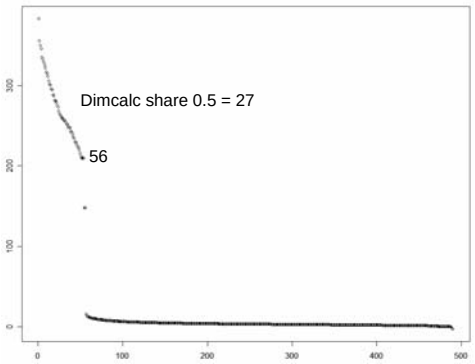
<49>

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

<50>

Singular Values of the Corpus



<51>

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)

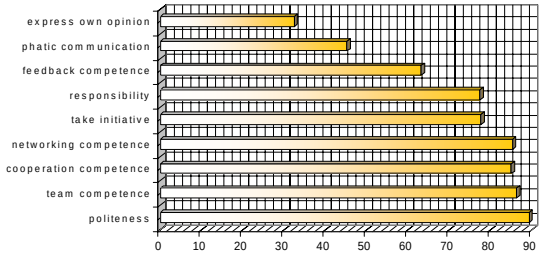
  = sum of r's = **2,7826994**
- polite = 0 for the documents
  - D311.txt (r = 0,6581647)
  - D126.txt (r = 0,6147096)
  - D055.txt (r = 0,6057419)
  - D034.txt (r = 0,5638606)
  - D299.txt (r = 0,5589148)
  - D031.txt (r = 0,5246403)

  = sum of r's = **3,5260319**

=> document T011.txt is inferred to be **polite = 0**

<52>

Results (Percentage of Correct Classifications)



<53>

Topic-Based Group Formation

<54>

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

<55>

Topic-Based Group Formation

- Input documents:
  - Self-Description (CV, Abstract, Paper)
  - Cluster Descriptions
- Measure LSA-Similarity between Cluster and self-description document (cf. Wild, 2006)
- Take average of all self-descriptions
- Take best-matching cluster as recommendation

<56>

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

<57>