The Role of Information Scent in On-line Browsing:

Extensions of the ACT-R Utility and Concept Formation Mechanisms

Peter Pirolli User Interface Research Area

Supported in part by Office of Naval Research (ONR) Advanced Research and Development Agency (ARDA) Palo Alto Research Center

Information Scent





Browsing often requires that we use visual or textual cues to direct our actions.

These local (proximal) cues are called *information scent*.

The *theory of information scent* is a psychological theory

Used to develop new user interface designs and new Web site evaluation tools





(a)



(d)

(c)

Overview

- Why information scent is important
- SNIF-ACT Model of Navigation
 - Where to go next (navigational choices)?
 - When to stop (leaving a Web site)?
- InfoCLASS Model of topic category formation
 - What categories of topics are where (learning about the information environment)



Example: Looking for Pirolli's Personal Page on the PARC Site



Hierarchical Navigation Spaces



Strong Information Scent = Navigation Cost Linear in Depth

Weak Information Scent = Navigation Cost Exponential in Depth



Navigation Costs as Function of Information Scent



Probabilities within range of empirical values in Woodruff et al. (2002)





Phase Transition in Navigation Costs as Function of Information Scent









WWW Experiment & Model

- 12 Stanford University students
- 6 tasks
- 2 tasks analyzed and modeled for 4 participants
- Example tasks
 - Find specific movie posters for your new living room
 - Find dates for a performance by a comedy troupe

Pirolli, P. & Fu, W. (2003). SNIF-ACT: A model of information foraging on the World Wide Web. Proceedings of the Conference on User Modeling.

Web Behavior Graphs (WBGs)

Find a poster for the movie "Antz"



Find the tour schedule of the "Second City Comedy Troupe"



- Links providing better information scent yield more direct navigation
- People abandon Web sites when information scent diminishes

Card, S., Pirolli, P., Van Der Wege, M., Morrison, J., Reeder, R., Schraedley, P., & Boshart, J. (2001). *CHI Proceedings*

SNIF-ACT

- Scent-based Navigation and Information Foraging in the ACT theory
- Declarative knowledge
 - User goal (e.g. Find the poster for "Antz")
 - Perceived aspects of Web page & browser
 - Large spreading activation network representing word associations
- Procedural knowledge
 - Productions representing basic Web browsing actions
- Utility: Information Scent
 - Mutual relevance between link text and user goal



Cognitive Model of Information Scent



Spreading activation





Base-level reflects log odds of occurrence

$$B_i = \ln(\frac{\Pr(i)}{\Pr(not i)} \psi)$$

Strength of link spread reflects log likelihood odds of cooccurrance

$$S_{jj} = \ln(\frac{\Pr(j|i)}{\Pr(j|not i)}\psi)$$

spreading activation networks



Information Scent: Random Utility Model based on Activation





Evaluation:Link-following actions



Observed vs Predicted Link Choice



Theory: Decision to Leave Web Site



Data: Sequences of Moves Just Prior to Leaving a Site





Match of SNIF-ACT to User Data







Forming Concepts about Information Sources

Scatter/Gather Document Clustering Browser

Scatter/Gather		
Text Database: trec1 This Window: 0 Parent Window: -1 Statter/Gather Show Titles	Start New Query OUIT Scatter/Gather	
🗌 Cluster 0 (38940)	☐ Cluster 5 (122371)	
cell, patient, radiation, dose, bean, disease, treatment, ima MP: Early Results In Nospital Patient Study Sho (aid, study, percent, DOE: Doses of secondary radiation appearing as (radiation, dose, expo AP: Poll: AIDS Test Confidentiality Opposed in (percent, study, drug,	user, software, computer, network, ibm, version, pc, line, d ZF: Hetworks for the '30s: getting down to busi (user, software, appl. ZF: Hindows 3.0 wins with the users. (News fmal (application, user, w ZF: Visual programning: is a picture worth 1.00 (software, program, u	
🗌 Ouster 1 (23564)	Cluster 6 (46144)	
court, judge, law, attorney, appeal, lawyer, trial, justice, MP800502-0111 MSJ: Supreme Court to Review New York Law That (court, state, u.s., f MSJ: Law - Legal Beat: Court Affires Vordict i (court, case, state,)	section, 1988, 1989, public, regulation, office, action, agen FR: Subpart 1852, Linstructions for Using Provi (1, section, informat FR: Community Bevelopment Block Grants (section, 1, state, r FR: Proprietary Trading Systems (commission, file, 19	
🗌 Cluster 2 (37760)	☐ Cluster 7 (48384)	
Fire, nile, spokesman, monday, friday, thursday, sunday, wedn MP: Troope Violated Rules in Village Raid Last (police, kill, report, MP: Up 10 50000 Reported Killed In Armenian Ou (official, report, gov AP: Two Palestinians Hounded in Clash Mith Sold (police, kill, army, a	0, phase, reaction, concentration, 6, cm, investigate, metal, DOC: The composition of MO/sub x/ generated in (2, 3, 1, rate, react DOE: Using a 40 cn bent crystal spectrograph of (2, 1, 3, refs, pas, DOE: X-ray method is used to study phase compos (2, 4, 3, 5, solution	
□ Cluster 3 (106437)	□ Cluster 8 (116210)	
american, bush, country, house, soviet, party, leader, think MP; U.S. Invasion of Panama Draws Fire From Lat (state, government, u. HSJ: Prosident's Pouce I Silpiping: Soviet (u.s., government, sov HSJ: The Iraqi Invasion of Kuwait Persian Gu (u.s., soviet, governm	coal, fuel, flow, technique, oil, pressure, production, mean DOE: A specialized software system which allowe (data, time, measurem DOE: No present results of our theoretical and (refs, tab, model, 1, DOE: Distributed control systems provide their (process, control, de	
Guster 4 (174353)	☐ Cluster 9 (28670)	
share, stock, trade, billion, bank, sale, sell, york, buy, e HSJ: Reliance Electric Mulle Public Offering: B (nillion, company, sto HSJ: Enterprise: 0 Staying Private Helped Rainb (company, nillion, sal MP: HMSIMWGTOW: he said. (company, nillion, sho	field, nuclear, plant, waste, reactor, magnetic, plasma, theo DOE: The effect of the radial electrostatic field, magnetic, pla DOE: The operational principle of nuclear power (plant, nuclear, anal DOE: Milson 'remormalization group' fixed point (field, theory, equat	





Browser Study

- Browsers
 - Scatter-Gather (N = 8)
 - Standard Similarity Search Engine (N = 8)
- 12 Tasks (TREC)
 - E.g., Find documents on new medical procedures for cancer

Pirolli, P., Hearst, M., Schank, P., & Diehl, C. (1996). Scatter-Gather browsing communicates the topic structure of a very large text collection. *Proceedings of the 'CHI '96 Conference*.



Inferred Topic Structure







Coherence of Mental Topics Among Users



Hierarchical clustering of users based on similarity of topic trees





InfoCLASS

- Information <u>Category Learning by Adaptation</u> to <u>Scent Stimuli</u>
 - Rational Analysis Model of Human Category Learning (Anderson, 1991, 1991)
- Assumes
 - Topics are mental categories
 - Mental categories are probabilistic collections of concepts
 - Links, thumbnails, citations provide information scent cues that evoke mental concepts



Evoking Categories and Inferences



share, stock, trade, billion, bank, sale, sell, york, buy, e HSJ: Relamce Electric Mulle Public Offering: 8 (million, company, stu BSJ: Enterprise 9 Sauging Private Helped Baib (company, million, sha DF: HMSHIDMETONF: ... he said, (company, million, sha DF: MSHIDMETONF: ... he said, (company, sillion, sha

Strength between observed UI features (words) and category reflects log odds

inferred features reflects log odds



Cues correspond to existing mental topics: Choose most activated (highest odds)





Otherwise create new category





InfoCLASS Simulation

- Scatter/Gather (SG): expose to *cluster summaries* from user logs
- Standard Search (SS)
 Expose to search result lists from user logs
- Evaluate ADA (category coherence) for the two groups of users



Simulation Performance

Categories

			Cure	
	SG	SS	SG	SS
1	330	464	59	28
2	470	767	149	39
3	526	1926	142	33
4	579	1481	101	28
5	770	1331	120	28
6	420	565	152	45
7	370	963	151	16
8	507	1486	124	14
	496	1123	125	29

Items

Scatter-Gather (SG) simulations develop more categories than Standard Search (SS).

Same trend as Pirolli et al (1996) comparison of SG vs SS users.



Category structure and coherence



	ADA	x 10 ⁻⁶ bits
SG	.949	
SS	1.292	

Scatter-Gather simulations exhibit greater category coherence (less divergence).



Summary

- Importance
 - Quality of information scent can effect qualitative changes in navigation costs (linear to exponential)
 - Browsers can differ in the how they communicate the topic structure of a collection of information
- SNIF-ACT model of navigation
 - Navigation & when to stop
- InfoClass
 - Model of formation of mental categories of topic structure
- Applications
 - New Uls (ScentTrails, Relevance Enhanced Thumbnails, ...)
 - Web Site Evaluation tools (Bloodhound, Lumberjack)



BACKUP SLIDES



Eye Movements: High Information Scent (Pirolli, Card, & Van Der Wege, 2003, TOCHI)





Eye Movements: Low Information Scent (Pirolli, Card, & Van Der Wege, 2003, TOCHI)

Hyperbolic Browser - mfcdemo . 8 × H ers Fruits Grains nuat @Parmeci Tre @Euglena Veg tables (BDian) flue-Gree Bacteria Vegetable @Amocha: wivir tavirus ®H₄ **VIII** Mineral ical Funai ntempor Future @The M ents Histori



Information Scent and the Cost of Navigation

(based on Hogg & Huberman, 1987)

- **D** = depth of search hierarchy
- *z* = average branching factor
- (1- q) = prob. of false alarm (Pr[FA])
- $\mu(q, z) = qz$ = average no. branches explored
- $\begin{array}{l} A(U,\,q,\,z) &= \text{average no. nodes explored within distance } U \\ &= (1 \mu(q,\,z)^{U\,+1})/(1 \mu(q,\,z)) \end{array}$
- N(D, z, q) = average no. nodes examined before desired goal found

$$= \left[\frac{(z-1)q}{2} \right]_{s=1}^{D-1} \sum_{s=1}^{D-1} A(s-1, q, z)$$



Prototype Formation

- If there are no existing categories, then create a new category (k_{New}) and assign instance P to the category, otherwise
- Determine the prob that the instance comes from a new category, $\Pr(k_{New}|P)$, and compare that to the existing category with the highest probability of including the instance, $\Pr(k_{Max}|P)$
 - Assign P to k_{New} if $Pr(k_{New}|P) > Pr(k_{Max}|P)$, else
 - Assign P to k_{Max}





$$\Pr(k_{New}) = \frac{1-c}{(1-c)+cn}$$



Categorization Model				
$W_1 \ W_2 \ \cdots \ W_N$ Item = [2 0 1 0 0]	Series of multinomial trials			
$\mathbf{W}_1 \ \mathbf{W}_2 \qquad \mathbf{W}_N$ Category = [.5 .1 .4 .1 .1]	Vector of probabilities			
$\Pr(\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_n \mid \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n) = \frac{1}{Z(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n)}$	$\overline{(\alpha,\alpha)}\prod_{i}p_{i}^{\alpha_{i}-1}$ Dirichlet			
Expected value	$\sum_{i} p_{i} = 1$ $\mathbf{E}[p_{i}] = \frac{\alpha_{i}}{\alpha_{0}}$ $\alpha_{0} = \sum_{i} \alpha_{i}$			
Noninformative prior	$\alpha_i = 1$			
Posterior distribution	$\mathbf{E}[\boldsymbol{p}_i] = \frac{\alpha_i + n_i}{\alpha_0 + \sum_j n_j}$			

Average Divergence from Average Entropy

$$D(p || q) = \sum_{x} p(x) \log \left(\frac{p(x)}{q(w)} \right)$$
$$ADA = \frac{\sum_{k=1}^{K} D(\mathbf{p}_{k} || \overline{\mathbf{p}})}{K}$$





Category structure and coherence



