Comparison Shopping Agent Systems- ShopBot

2000. 5. 17.
99419-514
박상욱

Question

- Explain the learning phase of ShopBot.

Contents

- A scalable comparison-shopping agent for the World-Wide Web Autonomous Agents 97, Robert B. Doorenbos, Oren Etzioni, and Daniel S. Weld
  - Intelligent Web agents
  - Online shopping agents
  - ShopBot
- A more scalable comparison shopping agent Engineering of Intelligent Systems(EIS 2000), HCI2000 J. Yang, J. Choi, J. Kim, H. Ham, K. Lee

Intelligent Web agents

- Design issue
  - Ability: To what extent can intelligent agents understand information published at Web site?
  - Utility: Is the agent's understanding sufficient to provided genuinely useful assistance to users?
  - Scalability: Can the agent automatically extract information from unfamiliar Web sites?
  - Environmental Constraint: What properties of Web sites underlie the agent's competence?

Online shopping agents

- Requirements of shopping agents
  - Help the user decide what product to buy.
  - Find specifications and reviews of the products.
  - Make recommendations.
  - **Comparison shopping to find the best price for the desired product.**
  - Monitor "What's new" lists and other sources to discover new relevant online information sources.
  - Watch for special offers and discounts.

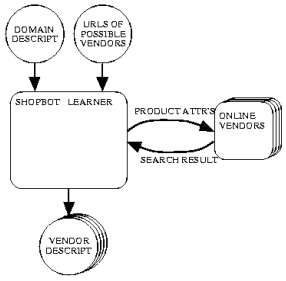
ShopBot(1)

- About ShopBot
  - Fully-implemented, domain-independent comparison-shopping agent.
  - Given the home pages of several online stores, autonomously learn how to shop at those vendors.
  - After learning, it is able to speedily visit shops, extract product information, and summarize the results for the users.
  - Require only minimal knowledge about different product domains without natural language processing.
  - Heuristic search, pattern matching, inductive learning techniques.

ShopBot(2)

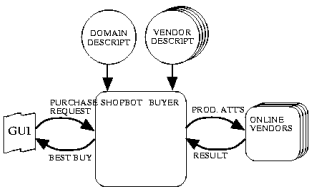
- Operate in 2 phases
  - Learning phase
    - Offline learner creates a vendor description for each merchant.
  - Comparison-shopping phase
    - Real-time shopper uses these descriptions to help a person decide which store offers the best price for a given product.

ShopBot(3)



- Learning phase
  - Analyze online vendor sites to learn a symbolic description of each site.
  - Domain model
    - Example products
      - $P_1, P_2, \ldots, P_n$.
    - Attributes of the products
      - ex) manufacture(P_1)= MS, name(P_1) =Win95

ShopBot(4)



- Comparison-shopping phase
  - Use learned vendor descriptions to shop at each site and find the best price for a specific product.
  - Execute the extraction procedures found by the learner for a variety of vendors

ShopBot(5)

- Environmental Regularities
  - The navigation regularity
    - Online stores are designed so consumers can find things quickly. Ex)searchable index, form.
  - The uniformity regularity
    - Use a uniform look and feel.
  - The vertical separation regularity
- Online vendors obey these regularities because they facilitate sale to human users.

Create vendor description(1)

- Vendor description
  - The URL of a page containing a form for a searchable index.
  - A function mapping product attributes to fields of that form.
  - Functions for extracting product data from pages returned by the index:
    - Recognize failure pages
    - Strip header and tail information from successful pages.
    - Extract a set of individual product descriptions

Create vendor description(2)

- Problem of extracting product description
  - Web page typically contains not only one or more product descriptions, but also information about the store itself, meta-information about shopping process, headings, links to related sites, and advertisements.
  - Meta-information
    - "Your search for Windows matched 3 items"
    - "Your shopping basket is empty"
  - Unsupervised learning

Create vendor description(3)

- Problems of learning a vendor description
  - Identifying an appropriate search form
  - Determining how to fill in the form
  - Discerning the format of product descriptions in pages returned from the form
  - Interdependent
    - The learner cannot be sure that a certain search form is the appropriate one until it knows it can fill it in and understand the resulting pages.

Create vendor description(4)

- Overview
  - Find a set of candidate forms.
  - For each form F_i, compute an estimate E_i of how successful the comparison-shopping phase would be.
  - Determine how to fill in the form, then make several test queries using the form to search for several popular products.
    - Training examples for the product descriptions
    - Compute E_i

Create vendor description(5)

- Finding and analyzing candidate forms
  - Start at the vendor's home page and follows URL links, performing a heuristic search looking for any HTML forms.
  - Discard forms that are clearly not searchable indices. (user name, address, phone number)
  - Associate matching attribute with the field.
    - Domain description contains synonyms for each attribute.
    - It matches the text preceding a field.

ShopBot(11)
Identifying product description formats

- **To determine the format of product descriptions in pages returned from the form**
- 3 subproblems
  - Learn a generalized failure template
    - Queries for dummy products
    - Build a generalized failure template
  - Learn to strip out irrelevant header and tail information
  - Learn product description formats

Identifying Product description formats(2)

- Success page
  - Queries for several popular products
  - Test result page for failure template
  - If majority of result page is failure, the form is not appropriate search form for the vendor
  - Otherwise, learner records generalized templates for the header and tail of success pages

Identifying Product description formats(3)

- Extracting product description
  - Assume all product descriptions have the same format at a certain level of abstraction (uniformity regularity)
  - Abstract language: HTML tag + keyword *text*
  - Every product description starts on a fresh line.
    - Break body of each result page into logical lines
  - Heuristic ranking function
    - sum of the number of lines, the number of price was found , the number of required attributes found

Generation the Vendor Description

- Decide which form is the best one
  - Repeat previous process for every form.
  - Based on making an estimate E_i
  - E_i reflects both the number of the popular products that were found and the amount of information present about each one.
  - Once learner has chosen a form, it records a vendor description for future.

Real-Time Comparison Shopping

- The operation of the shopper
  - Once it has received a request from the user via the GUI, it goes in parallel to each online vendor's searchable index, and fills out and submits the forms.
  - Find successful page, and strip off the header and tail, find production description ,sorts the results by ascending order of price, and generate a summary for the user.

Empirical Results

- Evaluate ShopBot Utility
  - Experiments
    - 3 groups
      - Those who used ShopBot (3)
      - Those who used Netscape's search tools and were also given the URLs of 12 software stores used by ShopBot(2)
      - Those who were limited to Netscape's search tools(2)

Group	Time (min:sec)	Navigator	eXceed	Word	Quicken
1	13:20	$30.71	$373.06	$282.71	$ 42.95
2	112:30	38.21	(not found)	282.71	41.50
3	58:30	40.95	610.00	294.97	42.95

Table 3: Subjects using the ShopBot performed the task much faster and generally found lower prices.

Empirical Results(2)

- Acquisition of new software vendors

Home Page URL	Navigator	eXceed	Word	Quicken
http://www.internet.net/	$ 28.57		$ 282.71	$ 43.06
http://www.cybout.com/cyberian.html	36.95		289.95	42.95
http://nscxdirect.necx.com/	31.95		329.95	42.95
http://www.sparco.com/	35.00		312.00	49.00
http://www.warehouse.com/	39.95		-	-
http://www.cexpress.com/	?	?	-	?
http://www.avalon.nf.ca/	44.95			
http://www.azteq.com/	?		?	?
http://www.cdw.com/			289.52	-
http://www.insight.com/web/zdad.html			315.00	
http://www.applied-computer.com/twelcome.html		$ 349.56		43.47
http://www.sidea.com/	59.00			

Table 4: Prices found by ShopBot for the four test products at twelve software stores. A space left blank indicates that ShopBot successfully recognized that the vendor was not selling this product; "?" indicates ShopBot found the product but did not determine the price; "-" indicates that ShopBot failed to find the product even though the vendor was selling it.

Empirical Results(3)

- Generality across product domains
  - CD domain (8 shops)
  - 1 day's work on describing the CD domain
  - Shop successfully at 4 CD stores
  - BargainFinder(hand-crafted) currently shops successfully at 3 stores

Conclusion

- ShopBot
  - Advantage
    - Find better prices in dramatically less time
    - Scale to multiple stores and product domains
  - Disadvantage
    - Not able to distinguish between upgrades to a product and product itself - need to do a more detailed analysis of product descriptions
    - Relies on a very strong bias

Conclusion(2)

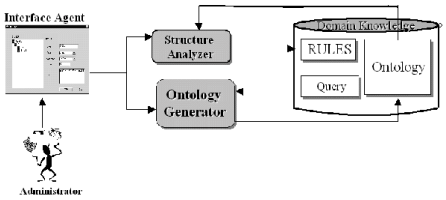
- Future works
  - Shopbot needs to be able to navigate a hierarchical organization.
  - Shopbot should be able to run Java applets and attempt to analyze their output

Appendix

- The Shopbot Research Prototype was retired in 1998.
  - It was a WWW shopping agent that enabled you to shop for computer software or CD's by manufacturer or title and artist queries.
  - This technology was adapted in private industry and contributed to the development of an even better shopping agent at www.jango.excite.com.

Automatic ontology generator

- Ontology
  - A set of some terminology or keywords



Automatic ontology generator(2)

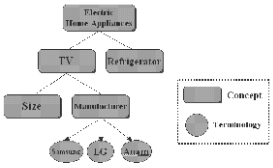


Figure 9: The structure of ontology

1: Ontology interface agent sends several different queries

2: Extract information

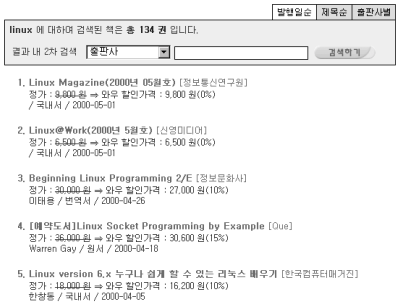
3: Extract terms and values from the PDU

4: Add the term to the ontology that has the most similar value to that in the existing ontology

Learning mechanism for tables

상 품 명	제 조 사	판 매 가	마일리지	비교
커천에이드 냉장고 4KSRS270F (749L)	Kitchen Aid	2950000원	0원	
LG 냉장고 R-B310DJ (314L)	LG전자	424000원	0원	
대우 냉장고 FRB-5280KB (520L)	대우전자	789000원	0원	
LG 냉장고 R-B230DJ (230L)	LG전자	311000원	0원	
[신제품]삼성 냉장고 SR-5549M (554L)	삼성전자	831300원	0원	
GE 냉장고 TFK28PFB 793L	GE	4150000원	0원	
LG 냉장고 R-B550MJ (550L)	LG전자	822000원	20000원	
GE 냉장고 22PRB (622L)	GE	2380000원	0원	
[신제품]삼성 냉장고 SR-5193M (514L)	삼성전자	766000원	0원	
[출시기념초특가]삼성 냉장고 SR-5190M (514L)	삼성전자	718000원	0원	
[신제품]삼성 냉장고 SR-2719D (265L)	삼성전자	377700원	0원	
GE 냉장고 TFK25ZAC 709L	GE	2100000원	0원	
LG 냉장고 R-B140DJ (137L)	LG전자	251000원	0원	
[신제품]삼성 냉장고 SR-5139M (514L)	삼성전자	816800원	0원	
LG 냉장고 R-A05DDJ (46L)	LG전자	129000원	0원	

Learning mechanism for lists



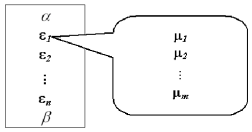


Figure 5: The general form of list-type results

α: header, β: tail

ε: product description unit

μ: each attribute

We have to find ε in result page

Step 1: Save the resulting page for the sample query into a file

Step 2: The page is broken down into some logical lines

Step 3: Analyze each logical line to recognize its meaning and expresses it by a number

Step 4: Entire page is expressed by a sequence of numbers. Then find a repeated pattern in this sequence.

<a href=CGI address>Bioinformatics: The ⋯	0
<dd>	2
Pierre Baldi, Soren Brunak / Hardcover ⋯	9
 	5
Our Price: $44.00	1
 	5
Average Customer Review ⋯	9
 	5
<a href=CGI address>Read more about ⋯	9
 	5

Figure 7: An HTML source and the pattern