ORIGINAL PAPER

Critiquing-based recommenders: survey and emerging trends

Li Chen · Pearl Pu

Received: 8 September 2010 / Accepted in revised form: 20 May 2011 / Published online: 19 October 2011 © Springer Science+Business Media B.V. 2011

Abstract Critiquing-based recommender systems elicit users' feedback, called critiques, which they made on the recommended items. This conversational style of interaction is in contract to the standard model where users receive recommendations in a single interaction. Through the use of the critiquing feedback, the recommender systems are able to more accurately learn the users' profiles, and therefore suggest better recommendations in the subsequent rounds. Critiquing-based recommenders have been widely studied in knowledge-, content-, and preference-based recommenders and are beginning to be tried in several online websites, such as MovieLens. This article examines the motivation and development of the subject area, and offers a detailed survey of the state of the art concerning the design of critiquing interfaces and development of algorithms for critiquing generation. With the help of categorization analysis, the survey reveals three principal branches of critiquing based recommender systems, using respectively natural language based, system-suggested, and user-initiated critiques. Representative example systems will be presented and analyzed for each branch, and their respective pros and cons will be discussed. Subsequently, a hybrid framework is developed to unify the advantages of different methods and overcome their respective limitations. Empirical findings from user studies are further presented, indicating how hybrid critiquing supports could effectively enable endusers to achieve more confident decisions. Finally, the article will point out several future trends to boost the advance of critiquing-based recommenders.

L. Chen (🖂)

P. Pu

Department of Computer Science, Hong Kong Baptist University, Hong Kong, China e-mail: lichen@comp.hkbu.edu.hk

Human Computer Interaction Group, School of Computer and Communication Sciences, Swiss Federal Institute of Technology in Lausanne (EPFL), Lausanne 1015, Switzerland e-mail: pearl.pu@epfl.ch

Keywords Critiquing-based recommenders \cdot Survey \cdot Preference elicitation \cdot Example critiquing \cdot Dynamic critiquing \cdot Hybrid critiquing \cdot User evaluations

1 Introduction

As online product catalogs have evolved to a second generation where users increasingly search and buy high-risk products such as apartments, laptops or cameras in online environments, the task of locating a desired choice among a large set of options is indeed becoming intimidating for the average customer. Recommender technology can help users find ideal products by exploiting user similarity (Konstan and Riedl 2012). So far collaborative filtering techniques have been used to suggest highly rated items from users who have experienced the products to similar users. However, for high-risk product domains, users are likely to search and buy products for the first time. This means the system cannot establish a meaningful profile for many of its recommendation seekers. To overcome such cold start problems, critiquing-based recommender systems have emerged and been broadly recognized as an effective preference based search and recommender technology, employing a distinct feedback mechanism called critiquing. In critiquing based systems, the initial user profile model does not influence the accuracy of the decision. Rather, it is the subsequent process of incremental critiquing that assists users in making more informed and accurate decisions.

Essentially, the critiquing-based recommender system simulates an artificial salesperson which will first recommend options based on users' current preferences and then elicit users' feedback in form of critiques (such as "*I would like some-thing cheaper*" or "*with faster processor speed*" when the product is Laptop). For a user to finally reach her ideal product, a number of such cycles are often required. Indeed, users are likely to construct their preferences in a context-dependent and adaptive fashion during the decision process (Tversky and Simonson 1993; Payne et al. 1993, 1999; Carenini and Poole 2000). A typical buyer has many constraints and preferences that are not stated up front. He or she becomes aware of these latent preferences only when proposed solutions violate them (Pu and Faltings 2000, 2002).

Critiquing systems help users incrementally build their preference models and refine them as they see more options. Consequently users are able to make more accurate and confident decisions even in complex decision environments. These systems have also been termed "conversational recommender systems", "conversational case-based reasoning systems", and "knowledge-based recommender systems" (Smyth and McGinty 2003; Shimazu 2001; Burke et al. 1997; Burke 2000). The system's main component is thus the *recommendation-and-critiquing*, which has been variously named *tweaking* (Burke et al. 1997), *candidate/critiquing* (Linden et al. 1997), *criteria editing* (Pu and Faltings 2000; Jurca 2000) and later *example critiquing* (Pu and Faltings 2004), *profiling by examples* (Shearin and Lieberman 2001), *navigation by proposing* (Shimazu 2001), and *critiquing feedback* (Smyth and McGinty 2003).

Typically, a critiquing-based recommender system follows the system-user interaction model as depicted in Fig. 1.



Fig. 1 The typical interaction model between users and a critiquing-based recommender system

- Step 1: the user is asked to first specify a reference product as the starting point; alternatively, she or he can give some specific value preferences on product features (e.g., their criteria on the apartment's price, size, etc.);
- Step 2: the system then returns one or multiple recommended items according to the user's initial preferences;
- Step 3: at this point, the user can select an item as her final choice and terminate the interaction. However, in most cases, a user is likely to choose a nearly desirable item and make critiques on the item. The specific types of critiques she makes depend on the system's critiquing support. As we will discuss later, critiquing supports fall into three major types: natural conversational critiques, system-suggested critiques, or user-initiated critiquing. The latter two methods have been more popularly deployed in existing systems;
- Step 4: once the critiques are made, the system will update its recommendations and return them in the next interaction cycle.

This interaction process continues until the user decides that she has found her most preferred product.

As an example, we consider a simple apartment finder interface that illustrates how critiquing-based recommenders generally work (Figs. 2, 3, 4). In Fig. 2, a user starts the search by specifying one or several preferences in the "search panel" area. Based on this initial preference model, the search engine will find and display a set of matching results. For example, seven matching apartments were displayed in this interface (Faltings et al. 2004a,b).

When a user is ready to select an apartment to put in the basket, the critiquing interface will be shown (see Fig. 3) where he/she can compare his/her current selection with others and perform critiquing analysis. For instance, suppose that the current selection is apartment 34. In the comparison window, the user can specify his/her desire for a bigger apartment by clicking on the checkbox next to the "bigger area" label. However, knowing that he/she may sacrifice something for a bigger apartment, he/she specifies "compromise" for both distance and kitchen attributes. Compromise means that a user is willing to accept a less satisfied value of the respective attribute.

Once a set of critiques has been composed, the system will show another set of matching examples (see Fig. 4). At this moment, apartment 31 seems quite interesting: it is around the same price and is 5 square meters bigger; however, it requires 10

					Apartm	entFinder - Exa	ample C	ritiquing In	lerface				
Searc	h Panel												
				impo	rtance					i	nportance		
tep	-			least	mo	ist		C.		least	most		
T	ype	room in a house	•	0.9	010	Bathroo	m	private	•		010		
P	rice Max	600	¥ Fs		010	Kitchen		private	•		.010	Show Results	s
	roa Min	20	• m2	0.0	010	Distanc	o May	10			0.010		
	a co mai	2.0	·			- Cristanc	C Mus	10			0.01.0		_
Searc	h Results												
F=									-				
	ID	type		price		area(m2)	ba	throom	1	dtchen	distance(min	15)	
tep4	4	room in a hous	e 600	563	15	10.00	private		private	8	15		
27		shared apartm	e 550	2	20		private	1	not av	ailable	10	Comp	are
33		shared apartm	e 600		18		share	a	not av	allable	5		
35		room in a nous	e 600		10		snare	a	private	3	20		
38		room in a hous	e 650		20		snare	a	snare	a	5	-	
42		shared apartm	8 700		25	1	private		private	3	10	Memo	rize
32		snareo aparen	e 600		114		snare	u	snare	a	10		
Baske	t ID	type		price	T.	area(m2)	bi	throom	[])	dtchen	distance(mir	15)	
34	1.1.10	room in a hous	e 600	1	15		private		private	3	15	-	
												Comp	ara
												12020	
												- 5000	100
												Finis	sh

Fig. 2 System showing a set of 7 results after a user's query

more minutes of commuting time and the bathroom is shared. This example illustrates that the system does not resolve the tradeoffs for the user; it rather provides relevant information for him/her to understand the decision context. The final choice is left to the user.

This *recommendation-and-critiquing* hence completes one cycle of interaction, which can continue as long as the user desires more of the critiquing process.

In the last decade, as more researchers have become interested in this area, three types of critiquing methods have been established: (1) *natural language dialog based critiquing support*, which acts as an artificial sales agent to communicate with the customer in a dialog interface, such as ExpertClerk (Shimazu 2001) and Adaptive Place Advisor (Thompson et al. 2004); (2) *system-suggested critiquing*, which proposes a set of critique suggestions for users to select, such as Dynamic Critiquing

(Reilly et al. 2004) and MAUT-based Compound Critiques (Zhang and Pu 2006); and (3) *user-initiated critiquing*, which aims at providing facilities to stimulate users to create critiquing criteria on their own, such as Example Critiquing (Chen and Pu 2006) and Flat Finder (Viappiani et al. 2006).

In this article, we would like to provide a survey of the start-of-the-art of critiquing systems as well as indicating the emerging trends for the future. We organize the rest of this article as follows: in Sect. 2, we will describe several recent and reprehensive critiquing-based recommender systems in the three respective branches; in Sect. 3, a hybrid framework will be presented, which aims at unifying the advantages of different methods and overcoming their respective limitations. Empirical findings from user studies will follow, indicating how the hybrid critiquing supports can effectively lead



Fig. 3 Critiquing support to guide users to critique the current example product for comparing it with the other tradeoff alternatives

end-users to achieve more confident and accurate decisions. Finally, via the development and analysis of taxonomy of the existing critiquing methods, we are able to indicate several future research and deployment directions in this area (Sects. 4-6).

2 Three types of critiquing systems

2.1 Natural language dialog based systems

The natural language method engages users in a conversational dialog and prompts them to provide preference feedback to the current recommendation. Two systems can illustrate how this process works.

2.1.1 ExpertClerk (Shimazu 2001)

ExpertClerk is an agent system imitating a human salesclerk. It interacts with shoppers in natural language and narrows down the set of matching goods by asking effective questions (*Navigation by Asking*). Then it shows three contrasting samples with explanation of their selling points (*Navigation by Proposing*) and observes the shopper's reaction. This process repeats until the shopper finds an appropriate good. Thus, the interaction obeys the critiquing model.

				Ара	artme	ntFinder - Examp	ole Critiquing In	lerface			
mpare											
			i i	nportance						importance	
			leas	t r	nost				lea	st most	
Туре	room in a house	•		001	۲	Bathroom	private	-	C	0010	
Price Max	600	-	Fs O	0.01	۲	Kitchen	compro	mise 🔻		0010	Show Results
Aros Min	himmer		m2 ()	0.01		Distance May	compromico		ine (0010	
Al Ca Milli	nigger	-	1112 O	100 100 1		Unstance may	compromise		iiiiə 🖉	0.010	
arch Resul	ts.										
ep 4	17.										
ID	type		lian	price	1	area(m2)	bathroom	kitch	en	distance(mins)	
34	room in a h	ouse	8 600		15	p	rivate	private		15	1
30	room in a h	ouse	550		20	p	rivate	private		35	Compare
31	room in a h	ouse	\$ 590		20	s	hared	shared		25	
35	room in a h	ouse	8 600		15	S	hared	private		20	
37	shared apa	intme	620		15	p	rivate	private		20	and a more than the
33	shared apa	irtme	600		18	S	hared	not availat	ole	5	Memorize
29	room in a h	ouse	9 550		22	S	hared	not availat	ble	20	
											-
ekot											
ISNEL											
ID	type		1 2	price	T	area(m2)	bathroom	kitch	en	distance(mins)	1
34	room in a h	ouse	600		15	p	rivate	private		15	1
31	room in a h	ouse	9 590		20	s	hared	shared		25	Compare
											Delete
											(Painters)

Fig. 4 System showing a new set of alternatives after the user's critiques

More specifically, after getting users' initial preferences via the conversational dialog, the system translates them into a SQL query and passes it to the database. If too many matching goods exist, the Navigation by Asking function would calculate the information gain of possible questions and then ask appropriate questions to the shopper for narrowing down the matching goods. After merchandise records are narrowed down to a pre-defined threshold number, the Navigation by Proposing function will show three significantly different samples. The first sample good is the good record closest to the center point of all matching goods. Its selling points directly reflect the customer's request. The second sample good is the record positioned farthest away from the center point, and the third sample good is the one positioned farthest away from the second sample. The explanation of the sample's selling point is given like "this is twice as expensive as those because it is made of silk and the other two are made of polyester". While seeing the explanation, the shopper could easily exclude one of the three proposed goods with a specific reason, such as "this one is too dark for me compared with the other two". The system will then modify the sample picking strategy accordingly.

2.1.2 Adaptive place advisor (Thompson et al. 2004)

This system employs natural language dialog for giving personalized place recommendations. Consider the following conversation between the inquirer (i.e., the user) and the advisor (i.e., the system):

- 1. Inquirer: Where do you think I should eat tonight?
- 2. Advisor: What type of food would you like?
- 3. Inquirer: What types are there?
- 4. Advisor: You can say things like Chinese, Indian, and Mediterranean.
- 5. Inquirer: Oh, maybe a cheap Indian place.
- 6.

To find items to recommend to the user, the Place Advisor carries out an augmented interactive constraint-satisfaction search. The goal of the entire conversation is to present an item that will be acceptable to the user. A probabilistic representation of the user's preferences, i.e., the query, is expanded considering both the long-term (over many conversations) and short-term (within a conversation). The system incrementally refines this query in the course of the conversation with the user, based on the user's critiques to the attributes and items offered during a conversation.

2.1.3 Discussion: pros and cons

The natural language dialog approach is especially suitable for generating recommendations that need to be delivered by speech interfaces, rather than visual platforms (e.g., while the inquirer is driving, etc.). This system also seems ideal, independent of modality, for practical tasks like destination selection or help-desk support, for which users need to converge on a few items.

However, since dialog interaction models demand precise natural language processing and high involvement of users, they are not so ideal to help an online user when they search for a product to buy in the e-commerce environment. Thus, some researchers have been engaged in the design and development of interactive critiquing in graphical user interfaces in order to better facilitate the process in such environments. These research works have resulted in two major types of approaches: *system-suggested critiques* and *user-initiated critiquing*.

2.2 System-suggested critiquing systems

The system-suggested critiquing system has been developed mainly to pro-actively generate a set of knowledge-based critiques that users might accept as ways to improve the current recommendation. The system produces the critiques according to its knowledge of the product domain, users' current preferences, or the availability of remaining products. For example, one of the earlier system-suggested critiquing systems was called FindMe. It helps users search through a large multi-dimensional information space using its knowledge about the product domain (Burke et al. 1996, 1997). A distinct interaction element in FindMe is its *tweaking* feature which allows users to critique the current recommendation by selecting one of the system's pre-designed tweaks (e.g., "cheaper", "bigger" and "nicer"). When a user responds to a tweak, the remaining candidates will be filtered to leave only those satisfying the tweak. Burke (2000) provides a detailed review of the FindMe systems. Another earlier example is ATA (Automated Travel Agent) developed by Linden et al. (1997). Its main aim was to help users select the ideal solution among configurable products, such as flights.

The main interaction element of ATA is called candidate/critique, which proposes a candidate solution and several critiques (e.g., "the best non-stop trip"). Through the critiques, the system elicits additional preferences from the user and then proposes better candidates in the next interaction cycle.

In recent years, the system-suggested critiquing mechanism has been largely improved by researchers in order to make the critiques more dynamically generated and being more accurately matched to users' needs.

2.2.1 Dynamic critiquing (McCarthy et al. 2004a, 2005c)

The motivation of developing dynamic critiquing was because of the limitations of FindMe system (Reilly et al. 2004): (1) its critique suggestions are pre-designed and fixed within a user's whole interaction session, so they are unable to reflect the user's changing needs and the status of remaining available products; (2) each of its critiques can only constrain over a single feature at a time (so called the *unit critique*). Therefore, Reilly et al. (2004) have proposed an alternative strategy with the aim to suggest a set of *dynamic compound critiques*, each of which is essentially a combination of multiple unit critique is "Different Manufacture, Lower Processor Speed and Cheaper"). Concretely, the Dynamic Critiquing interface (see Fig. 5) presents both unit and compound critiques to users as feedback options.

The system is essentially based on the association rule mining technique to discover frequent sets of value differences between the current recommendation and the remaining un-recommended products. The Apriori algorithm, a broadly applied association rule mining tool (Agrawal et al. 1993), was chosen to fulfill the task. More specifically, they use the Apriori algorithm to discover highly recurring compound critiques that are representative of a given dataset. They then filter all possible compound critiques by using a threshold value, favoring those critiques with lower support values ("support value" refers to the percentage of products that satisfy the critique). Such selection criterion was motivated by the consideration that presenting critiques with lower support values provides a good balance between their likely applicability to the user and their ability to narrow the search. Once the user selects a critique, a product satisfying the chosen critique as well as being most similar to the current recommendation is returned as a new recommendation in the next cycle.

A live-user trial (with 38 subjects) showed that the total number of recommendation cycles decreases from 29 to 6 when users actively selected the compound critiques (McCarthy et al. 2005b,c). On the other hand, the proposed dynamic critiques can perform as explanations, exposing to users the recommendation opportunities in the remaining dataset (McCarthy et al. 2004b; Reilly et al. 2005a).

In their follow-up work, the system was extended to *Incremental Critiquing*, which particularly records the user's critiquing history (i.e., critiques that the user picked in the past cycles), and hence is able to avoid repeatedly endorsing attribute value(s) that the user already stated dislike (Reilly et al. 2005b). A user evaluation revealed that the extended version can save users' interaction cycles by up to 34%, in comparison with the original system (McCarthy et al. 2005a).

📔 Digital Camera:	1	Shop for: <u>Digital Cameras</u> , <u>Holidays</u> ,
Comment	Adjust your prefere product for you!	ances in Unit Critiques ight
	Manufacturer	H Cyron H
	Model	H EOS-3000
Item Found: CASE2 Specifications 6.3 Megapixel CMOS sensor 7-point wide-area AF High-performance DIGIC processor 100-1600 ISO specifications	Price (\$)	+ 871
	Format	H R H
	Resolution (M Pixels)	♦ 6.29
	Optical Zoom (X)	* 10.0
	Digital Zoom (X)	0.0
	Weight (grams)	÷ 645.0
ompatible with all Canon EF	Storage Type	X Compact Flash
tBridge, Canon Direct Print and	Storage Included (MB)	0.0
C required	We have more matching	ing products with the following.
Comment	1. Less Optical Zoom & M Storage Type (139)	More Digital Zoom & A Different
Compound	2. A Lower Resolution &	A Different Format & Cheaper (169)
Critiques	3. A Different Manufactur	arer & Less Optical Zoom & More Exercise

Fig. 5 The Dynamic Critiquing interface with system suggested compound critiques for users to select (McCarthy et al. 2005c)

2.2.2 MAUT-based compound critiques and visual critiquing (Zhang and Pu 2006)

However, the Dynamic-Critiquing method (including its extension) is still limited, in that it only reveals what the system can provide, but does not take into account users' interest in the suggested critiques. Given this limitation, Zhang and Pu (2006) have proposed an approach with the purpose of adapting the generation of compound critiques to user preferences. Formally, they model each user's preferences based on the multi-Attribute Utility Theory (MAUT), which is a theory taking into account of conflicting value preferences and producing a sore for each item to represent its overall satisfaction degree with the user preferences (Keeney and Raiffa 1976). Thus, during each recommendation cycle, according to the user's current preferences, the top *k* products with maximal MAUT values are retrieved. The ranked first item is then taken as the top candidate, and for each of the others, its detailed value differences from the top candidate will be presented as a compound critique (e.g., "same brand with lower price, but slower CPU speed, smaller screen, smaller memory and smaller hard disk").

Relative to Dynamic Critiquing methods, these MAUT-based compound critiques were proven with significantly higher recommendation quality, inferring that they can



Fig. 6 An example of the compound critique with its original textual interface (above) and its new visual interface design (below) (Zhang et al. 2008)

be better matching to the user's intended critiquing criteria (Zhang and Pu 2006; Reilly et al. 2007).

In their later work, Zhang et al. (2008) modified their critiquing interface by means of a visual design, where originally text-only critiques were replaced with meaningful icons. For example, "cheaper" was replaced with a value-augmented icon. The green border represents positive change and the down arrow represents the decreasing direction (see Fig. 6). This interface was favored by more users and stimulated them to more actively apply the suggested critiques.

2.2.3 Preference-based organization interface (Pu and Chen 2006; Chen and Pu 2007c)

Although MAUT-based compound critiques and its visual design were proven more effective than Dynamic Critiquing systems, it is still in nature limited, in that each MAUT-based compound critique corresponds to one product only and not much recommendation opportunities can be exposed to users. Moreover, each critique contains too many attributes which will likely cause information overload to the user.

Thus, preference-based organization technique was proposed to compensate for the above-mentioned limitations (Chen and Pu 2007c). It was designed not only dynamically generating critiques adaptive to users' MAUT based preference model, but also applying the association rule mining tool to discover compound critiques that can be representative of the remaining dataset. In addition, the critiques and their contained products are diversified so as to assist users in refining and accumulating their preferences more effectively. Briefly speaking, the algorithm contains three main steps (see details in (Chen and Pu 2007c): (1) it establishes the user's preference model by representing her initial preferences as a weighted additive sum of value functions according to MAUT; (2) it computes the recommendations based on the user's preference model, and each of them (except the top candidate) is converted into a tradeoff vector, comprising a set of (attribute, tradeoff) pairs; the tradeoff in essence indicates the *improved* or *compromised* property of the product's attribute value being compared to the same attribute of the top candidate; for the attributes without explicitly stated preferences, the default properties are suggested (e.g., the cheaper, the better); (3) it organizes the tradeoff vectors into categories through the Apriori algorithm (Agrawal et al. 1993). The algorithm helps discover the recurring subsets of (attribute, tradeoff) pairs as candidates of system-suggested critiques. Since a large amount of subsets are usually produced by this association rule mining tool, these candidates are further ranked and diversified, and the ones with higher tradeoff utilities (i.e., gains vs. losses relative to user preferences) are finally presented in the organization interface. In the interface, after a user selected one of the suggested critiques and furthermore a reference product from the set of recommendations that satisfy the selected critique, her preferences will be accordingly refined. Formally, the weight (i.e., the relative importance) of improved attribute(s) that appears in the selected critique will be increased, and the weight of compromised one(s) will be decreased. All attributes' preferred values will be also updated based on the selected product's values. This refined preference model will then be inputted to the recommendation engine to generate a new set of critique suggestions in the next cycle. This process continues till the user makes the final choice.

Figure 7 shows the sample interface, where multiple products that satisfy the suggested critique are recommended together. This interface was favored by most of the interviewed users since it could potentially save their interaction effort and give them high control over the process of making choice. A comparative study assessed its algorithm accuracy, by comparing it to FindMe, Dynamic Critiquing and MAUT-based compound critiques. The experimental results showed that it can achieve significantly higher critique prediction accuracy and recommendation accuracy. It was also proven to be most effective in terms of saving users' interaction effort.

2.2.4 Discussion: pros and cons

It can be seen that the pros of system-suggested critiques are mainly their ability of educating and guiding users to provide feedback criteria, so that the system can better understand users and enhance its recommendation power. It has also been proven that the critique suggestion accuracy can be highly improved if they take into account of users' current and potential interests. This infers that if the critiques optimally match to what users are prepared to make, the system will likely save users' interaction effort and accelerate their decision performance. On the other hand, the critique suggestions can perform as explanations and help users be familiar with the product domain and the relationship between attributes. Users can be then stimulated to express more preferences and/or be prevented from making retrieval failures (Reilly et al. 2005a).

Although the critique prediction accuracy was significantly improved in recent studies, the accuracy is still not so ideal (i.e., the best accuracy was 66.9%, as achieved by the preference-based organization technique (Chen and Pu 2007c)). It hence indicates that the critique suggestions are unable to precisely match to users' intended criteria all the time. In the failure situations, because users have no chance to build critiques on their own, they would have to take more effort in order to locating target choice. For example, consider a user who is looking for a digital camera with higher resolution and more optical zoom than the current recommended product. Suppose there is no suggested critique matching to this criterion. At this point, she may either sacrifice her requirement on resolution and optical zoom by selecting the current product, or be involved in longer interaction session by pursuing other ways to locate the desired product. In both cases, the user would perceive that the system is not so competent in helping her make quick and accurate decision.

The top ca	ndidate	according	g to you	r preferences					
Manufacturer	Price	MegaPixels	Optical zoom	Memory type	Flash memory	LCD screen size	Depth	Weight	
Canon	\$242.00	5.0 MP	Эн	CompactFlash Card	32 MB	1.8 in	1.37 in	8.3 oz	choose
We have n	nore pro	oducts wit	h the fo	llowing					
they are ch	eaper a	nd lighter,	but hay	e fewer megapi	xels				
Nikon	\$167.95	4 MP	3x	SD Memory Card	14 MB	1.8 in	1.4 in	4.6 oz	choose
Canon	\$230.00	4.1 MP	З×	CompactFlash Card	32 MB	1.5 in	1.09 in	6.53 oz	choose
Canon	\$180.00	3.3 MP	Зx	SD Memory Card	16 MB	2 in	0.83 in	4.06 oz	choose
Canon	\$219.18	4.2 MP	4×	MultiMedia Card	16 MB	1.8 in	1.51 in	6.35 oz	choose
Canon	\$163.50	3.2 MP	4×	MultiMedia Card	16 MB	1.8 in	1.5 in	6.3 oz	choose
Canon	\$199.40	3.2 MP	2.2x	SD Memory Card	16 MB	1.5 in	1.4 in	5.8 oz	choose
they have n	nore me	ganivele a	nd bigg	on coreene hut	Pe more	evnenciv			
Sony	\$365.00	7.2 MP	39	Internal Memory	32 MB	2.5 in	1.5 m	6.9.07	choose
Capon	\$439.99	7.1 MP	34	SD Memory Card	32 MB	2.0 11	1.04 in	6.02	choose
Euri	\$252.00	6.2 MP	- Au	XD-Richure Card	16 MB	2 10	1.4 in	71.02	choose
Sooy	\$236.00	7.2 MD	34	Internal Memory	22 MB	2 in	1 in	5.07	choose
Nikop	\$304.18	7.1 MD	34	Internal Memory	13.5 MB	2 in	14 m	5.3.07	choose
Olympus	\$334.00	7.4 MP	5×	XD-Picture Card	32 MB	2.0 in	1.7 in	7.1 oz	choose
		1 .1 .							
they are ng	nter and	a thinner,	bur nav	e less hash men	iory				
Pentax	\$238.99	5.3 MP	3x	Internal Memory	10 MB	1.8 in	0.8 in	3.7 oz	choose
Canon	\$273.18	4.0 MP	3x	SD Memory Card	16 MB	2 in	0.82 in	4.59 02	choose
Nikon	\$329.95	5.1 MP	Зx	Internal Memory	12 MB	2.5 in	0.8 in	4.2 02	choose
Canon	\$316.18	5.3 MP	3x	SD Memory Card	16 MB	2 in	0.81 in	4.59.02	choose
Casio	\$386.00	7.2 MP	3x	Internal Memory	8.3 MB	2.5 in	0.88 in	4.48 02	choose
Fuji	\$309.18	6.3 MP	Зx	XD-Picture Card	16 MB	2.5 in	1.1 in	5.5 02	choose
they have a	nore opt	tical zoom	with di	fferent memory	type, but	are thick	er and	heavi	Br
Panasonic	\$386.00	5.0 MP	12×	SD Memory Card	16 MB	1.8 in	3.34 in	11.52 02	choose
Konica Minolta	\$349.99	5.0 MP	12x	SD Memory Card	16 MB	2 in	3.3 in	12 oz	choose
Fuji	\$259.18	4.23 MP	10x	XD-Picture Card	16 MB	1.5 in	3.1 in	11.9 oz	choose
Olympus	\$253.00	4.0 MP	10x	XD-Picture Card	16 MB	1.8 in	2.7 in	9,9 02	choose
Olympus	\$284.99	4.0 MP	10x	XD-Picture Card	16 MB	1.8 in	2.7 in	10.6 02	choose
Nikon	\$259.18	4.2 MP	8.3x	Internal Memory	13.5 MB	1.8 in	2.2 in	9 oz	choose

Fig. 7 The preference-based organization interface where category titles behave as suggested compound critiques (Chen and Pu 2007c)

2.3 User-initiated critiquing systems

Instead of suggesting critiques for users to choose, the user-initiated critiquing approach focuses on showing examples and stimulating users to make self-motivated critiques. It does not limit the critiques a user can manipulate during each cycle, so that users can post either unit or compound critiques over any combination of features with freedom. In fact, the target of this kind of system is to assist users in freely executing tradeoff navigation, which is a process shown to significantly improve users' decision accuracy and confidence (Pu and Chen 2005).

Precisely, the tradeoff navigation involves finding products that have more optimal values on important attributes, while accepting compromised values for other less important ones. With the user-initiated critiquing interface, the user can conveniently start the tradeoff navigation from one item (called the reference product), specify her tradeoff criteria in terms of improvement and compromise regarding the product's attributes, and then see a new set of products that are nearly approaching to her ideal choice. The unit and compound critiques are respectively termed as simple and complex tradeoffs in such systems (Pu and Kumar 2004).

To find similar products wit	th better values that	n this one	
	Canon PowerS \$424.15 Canon, 5.3 M pixe 2.97 in thickness,	Shot S2 IS Digital Camer Is, 12x optical zoom, 16 MB r 404.7 g weight. <u>detail</u>	ra Add to saved list nemory, 1.8 in screen size,
would you like to improve s	ome values?		
	Keep	Improve	Take any suggestion
Manufacturer	Canon	O Sony	0
Price	○ \$424.15	Iess expensive	0
Resolution	• 5.3 M pixels	 \$100 cheaper 	0
Optical Zoom		\$300 cheaper	0
Removable Flash Memory	I6 MB	🔿 more memory 💌	0
LCD Screen Size		O larger	0
Thickness		thinner	0
Weight	⊙ 404.7 g	O lighter	0
	Show Re	sults	

Fig. 8 The Example Critiquing interface that turns the critiquing process into a decision tradeoff process (Chen and Pu 2006)

2.3.1 Example Critiquing (Pu et al. 2008)

The Example Critiquing agent is a purely user-initiated critiquing system. It originated from ATP, which was an online preference-based search tool for finding flights (Torrens et al. 1997).

The Example Critiquing system mainly consists of a user interface and a search engine. A user initially starts the search by specifying a few preferences in the query area. Each preference is composed of one acceptable attribute value and its corresponding degree of importance (i.e., weight). Behind, a MAUT-based preference model will be built. Based on the initial preference model, the search engine will rank all alternatives by their matching scores and return the top k ones (normally, k is between 5 and 20 according to (Faltings et al. 2004b) that suggests the optimal number). The user either accepts a result, or takes a near solution and activates the tradeoff navigation panel (i.e., the critiquing panel, see Fig. 8) where she can post critiques to the near solution. For instance, as shown in Fig. 8, three radio buttons are next to each attribute, respectively under "Keep" (default), "Improve" and "Take any suggestion" labels, thus facilitating users to critique one attribute by either improving its current value (i.e., selecting "Improve") or accepting a compromised value suggested by the system (i.e., via "Take any suggestion"). More notably, users can freely compose compound critiques by combining critiques on any two or more attributes.

Once a set of critiques is composed, the system will refine the user's preference model and adjust the relative weights of all critiqued attributes. Formally, the weight of improved attribute(s) will be increased and that of compromised attribute(s) will be decreased. The search engine will then compute and return a new set of recommendations based on the refined preference model.

Besides modeling user preferences based on MAUT, for configurable products (e.g., flight planning), the search engine employs sophisticated constraint satisfaction algorithm (Torrens et al. 2002). Formally, the question was defined as a constraint satisfaction problem (CSP) that includes both hard constraints (i.e., the criteria that should not be violated) and soft constraints. Solving a CSP involves finding solutions that are optimally preferred. For this purpose, the weighted CSP was utilized because it can find the optimal solution to minimize the weighted sum of preferences and considers tradeoffs.

Increasing users' decision accuracy has uniquely motivated Example Critiquing research in the past several years (Pu et al. 2011). In a set of more than 10 trials that involves real users, researchers showed that Example Critiquing systems were able to increase users' decision accuracy (Pu et al. 2008), save their cognitive decision effort (Chen and Pu 2009) and improve users' preference certainty and decision confidence in their search task (Pu and Chen 2005).

2.3.2 Flat Finder (Viappiani et al. 2007a)

Based on the Example Critiquing infrastructure, Viappiani et al. centered their research on how to help users express more preferences, by means of returning more predictive product recommendations. They developed Flat Finder which is based on the *lookahead* principle to provide product suggestions (Viappiani et al. 2006). The main idea is that, besides showing products that best match to the user's current preferences, it also returns products that will become optimal if the user adds a new preference. Technically, they considered Pareto-optimality to build a probability model for predicting users' new preferences: an option is *Pareto-optimal* if there is no other option that is better or equally preferred regarding all preferences, and strictly better w.r.t. at least one preference. Products are then evaluated according to their probability of becoming Pareto-optimal. Through series of simulation and user studies, they demonstrated that their method prompted users to articulate more preferences, and enabled users to achieve a higher level of decision accuracy (Viappiani et al. 2006, 2007a).

2.3.3 Discussion: pros and cons

The main advantage of user-initiated approach is that it allows for a higher level of user control. Users might feel at ease to create critiques in their own desires, whereas in the system-suggested system, they can only "select", but not "create". In addition, the system is more flexible because more critiquing facilities can be provided on the interface for users to specify their criteria. Prior works have also discovered these systems' positive effects on improving users' decision accuracy and quality, given their support to tradeoff navigation (Pu and Chen 2005).

However, users perceived less cognitive effort while using these systems, although their objective interaction effort is almost equivalent, and sometimes even higher, than their effort using the system suggested critiquing systems (Chen and Pu 2006). The previous user study also found that users normally required a warm-up period to be familiarized with this system, since these user-initiated critiquing facilities were not so easily grasped at their first glance. Thus, the main challenging issue for this type of system is how they could reduce users' effort consumption, while still maintaining high level of user control and decision accuracy.

2.4 Summary

In a summary, the systems described in this article can be generally classified in terms of their *critiquing supports: natural dialog* involves users in a conversational mode to state the criteria in natural languages, *system-suggested critiques* propose a set of critique options for users to select, and *user-initiated critiquing* stimulates users to freely create critiques on their own. Under this classification, we have further reviewed these systems from others aspects, such as the number of items that they recommended at a time for users to critique. As summarized in Table 1, each surveyed system is additionally characterized from the following dimensions:

Critiquing coverage concerns the number of recommendations that are returned in each cycle for users to provide feedback. Most of user-initiated critiquing systems, such as Example Critiquing (Pu and Chen 2005), returned multiple items during each recommendation cycle, so that users have the freedom to choose their desired critiquing object. On the contrary, system-suggested critiquing systems, including FindMe (Burke et al. 1997) and Dynamic Critiquing (Reilly et al. 2004), present one recommendation at a time and they base it to produce the suggested critiques. This simple display strategy should have the advantage of not overwhelming users with too much information, but it deprives users of the right to select the interested product, and hence potentially brings the risk of engaging users in longer interaction session.

Critiquing modality determines the type(s) of critiques that users can actually state to a product. Looking into above systems in this regard, we come up with three major types. The first one is similarity-based critiquing such as "Find a camera similar to this one." This type of feedback is also called preference-based feedback in (Smyth and McGinty 2003), and regarded as the least demanding approach w.r.t. user effort, domain expertise and interface complexity. The second is *quality-based critiquing* such as "Find a similar camera, but cheaper." This type of critiquing is suitable for users who desire improvement on specific attributes (e.g., price), but are unable to specify the exact amount to be improved. The third type is *quantity-based critiquing*, e.g., "Find something similar to this camera, but at least \$100 cheaper." It applies when users have exact value preferences, and this kind of critiquing can be efficient for users to filter out irrelevant items. Because the critiques from system-suggested critiquing systems are pre-formulated or dynamically generated, most of them support only quality-based critiques. They claimed that such critiques provide a compromise between the detail provided by value elicitation (i.e., quantity-based critiquing) and the ease of feedback associated with preference-based methods (i.e., similarity-based

	Critiquing generation	Critiquing coverage	Critiquing modality	Critiquing unit	Product cases
ExpertClerk (Shimazu 2001)	Dialog-based	Multiple products $(p=3)$			Clothes, wines
Adaptive Place Advisor (Thompson et al. 2004)	Dialog-based	One product			Restaurants
FindMe (Burke 2000)	System-suggested k critiques ($k = 7$ critiques in RentMe)	One product returned for users to critique	Quality-based	Unit critiques	Apartments, restaurants, cars
ATA (Linden et al. 1997)	System-suggested k critiques ($k = 2$)	Multiple products ($p = 5$) returned for users to select their critiquing object	Quality-based	Unit critiques	Flight tickets
Dynamic Critiquing (Reilly et al. 2004)	System-suggested k critiques ($k = 3$)	One product	Quality-based	Compound critiques	PCs, Digital cameras
MAUT-based Compound Critiques (Zhang and Pu 2006)	System-suggested k critiques ($k = 5$)	One product	Quality-based	Compound critiques	Apartments, PCs
Preference-based Organization (Chen and Pu 2007c)	System-suggested k critiques ($k = 4$)	Multiple products $(p = 6)$	Quality-based	Compound critiques	Laptops, digital cameras
Example Critiquing (Pu and Chen 2005)	User-initiated critiquing aid	Multiple products ($p = 7$)	Similarity-based; quality-based; quantity-based	Simple and complex tradeoffs (i.e., unit and compound critiques)	Apartments, laptops, digital cameras, flight planning
Flat Finder (Viappiani et al. 2006)	User-initiated critiquing aid	Multiple products ($p = 6$)		Simple and complex tradeoffs	Apartments

 Table 1
 Categorization of typical critiquing-based recommender systems

critiquing) (McCarthy et al. 2005c). In comparison, due to the flexibility of user-initiated critiquing interfaces, they are capable of facilitating different types of modalities

ated critiquing interfaces, they are capable of facilitating different types of modalities, especially the quantity-based ones. For example, in Example Critiquing interface (see Fig. 8), the option "\$100 cheaper" is listed in the pull-down menu of "price".

Critiquing unit refers to the number of product features that users are allowed to specify critiquing criteria each time. As noted before, system-suggested critiquing systems commonly classify their supports as either *unit critique* (i.e., the critique is made on only one feature, e.g., "cheaper" in FindMe), or *compound critique* (i.e., on multiple features simultaneously, such as MAUT-based compound critiques), or both (such as in Dynamic Critiquing). In Example Critiquing (Chen and Pu 2006), since it does not limit the type and unit of critiques a user can manipulate during each cycle, both unit critique and compound critique are supported in form of simple tradeoff and complex tradeoff respectively. For example, the user can improve or compromise a single feature and leave the others unchanged (i.e., unit critique), or combine multiple unit critiques into a compound critique.

3 Towards hybrid critiquing systems

As discussed before, each type of the systems inherently possesses certain pros and cons. In particular, it seems that the respective strengths of *system-suggested critiquing* and *user-initiated systems* could well compensate each other. Driven by this observation, a comparative user evaluation was conducted, which empirically revealed the two approaches' combined as well as individual merits. To follow-up, two variations of hybrid critiquing systems were developed.

3.1 User-initiated critiquing versus system-suggested critiques

In an experiment that involved 36 participants, two example systems, Example Critiquing (EC) and Dynamic Critiquing (DC) were compared, which respectively represented user-initiated and system-suggested critiquing systems. The results show that EC achieved significantly better results in terms of users' decision accuracy, cognitive effort perception and decision confidence (Chen and Pu 2006). Further analysis of users' written protocols uncovered that the primary factor that led to EC's success is its combination of multi-product critiquing coverage strategy and a user-initiated critiquing aid. This combination gave users a higher degree of control when they compared products and composed tradeoff criteria. However, some users (36.1%) still subjectively preferred the system-suggested DC, because they found it intuitive to use, and the critique suggestions motivated them to provide feedback to the current recommended product. Indeed, the majority of these 36.1% users were able to accelerate their decision process when the system suggested the critiques that they were prepared to accept.

This study motivated researchers to develop a new type of critiquing interface which could well combine both approaches' advantages, so as to maximally improve users' decision performance and subjective preference. The so called *hybrid critiquing-based recommender system* was hence proposed, and it is believed that in such system, people

To find similar products with	h better values than	this one			
io	Canon PowerS \$424.15 Canon, 5.3 M pixe 2.97 in thickness.	hot S2 IS Is, 12x optic 404.7 g weig	Digital Came al zoom, 16 MB ht. <u>detail</u>	ra Add to seved list memory, 1.8 in screen size,	The product bein critiqued
We have the following					
1. Less Optical Zoom and Thi	nner and Lighter Weig	lht	Explain	Show Products	System-suggester
2. Different Manufacturer and	Lower Resolution and	Cheaper	Explain	Show Products	compound critiqu
3. Larger Screen Size and Mo	re Memory and Heavi	er	Explain	Show Products	J
OR would you like to improv	ve some value(s) by	yourself?			
	Keep		Improve	Take any suggestion	\square
Manufacturer	 Canon 	() Son	· ·	0	
Price	\$424.15	O less	expensive 🛩	0	
Resolution	● 5.3 M pixels	O high	er 🖌	0	
Optical Zoom	● 12×	O more	zoom 🛩	0	User-initiated
Removable Flash Memory	16 MB	O more	memory 🛩	0	critiquing facility
LCD Screen Size	1.8 in		er 💌	0	
Thickness	② 2.97 in	O thinn	er 👻	0	
	C2003030	Concerne and		-	

Fig. 9 Hybrid critiquing system (version 1): the combination of system-suggested compound critiques and user-initiated critiquing facility (Chen and Pu 2007a)

can not only obtain knowledge of the domain and easily perform critiquing via the suggested critiques, but also have the opportunity to freely compose critiques on their own with the self-initiated critiquing support.

3.2 Hybrid Critiquing

Two hybrid critiquing systems have been developed so far. One was the combination of Example Critiquing facilities with Dynamic Critiquing based compound critiques on a single screen. The second version integrated the preference-based organization interface (Chen and Pu 2007c) (which shows system-suggested critiques and their associated products on a separate page) with Example Critiquing (the EC interface is evoked only when users activate it). Series of user studies were conducted on the two versions.

3.2.1 Version 1

Figure 9 shows a sample interface of the first type of hybrid critiquing system, where DC compound critiques are displayed with EC facilities on the same screen (Chen and Pu 2007a). Specifically, the current recommendation is displayed at the top and followed by multiple suggested critiques. The self-initiated critiquing area is placed below, which provides functions to facilitate various types of critiquing modality

(e.g., similarity-based, quality-based, or quantity-based) and critiquing unit (e.g., unit or compound critiques). In this interface, users can freely choose to pick the suggested compound critiques, or create their own critiquing criteria. For example, when a user is looking for some products with higher resolution and more optical zoom relative to the current recommended camera, if one of the suggested critiques exactly matches the requirements, she can undoubtedly select it. Otherwise, she can specify the criteria in the self-initiated critiquing panel (i.e., by "improving" the resolution and optical zoom simultaneously). Besides, she can also choose to compromise some of other attributes that are less important for her, so as to guarantee the intended gains.

After each critiquing process, a set of tradeoff alternatives that best match to the user's critiques will be returned by the system for users to compare. The search algorithm is accordingly chosen to adapt to the type of critiques users posted. For example, it applies similarity and compatibility selection measures if the DC-based critique is picked, and employs elimination-by-aspect (EBA) plus weighted additive sum rule (WADD) ranking mechanism if the user self specified critiques. Among the recommended items, if the user finds her target choice, she can proceed to check out. Otherwise, if she likes one product but wants something to be additionally improved, she can come back to the hybrid critiquing interface to resume a new critiquing cycle.

A user study (with 18 participants) studied users' actual behavior in this hybrid system (Chen and Pu 2007a). It showed that users behaved more actively in creating criteria with the self-initiated critiquing panel, relative to their application of the system-suggested critiques. Eventually, users obtained high levels of decision accuracy, decision confidence and behavioral intentions.

In addition, by comparing this hybrid critiquing system to systems without critique suggestions or without user-initiated critiquing facility, the two components' respective roles became much clearer. In fact, both significantly contributed to enhancing users' confidence in their decision and intention to return to the hybrid system for future use.

3.2.2 Version 2

As noted in Sect. 2.2, in comparison with Dynamic Critiquing and MAUT-based compound critiques that generated critique suggestions, the *preference-based organization* (*Pref-ORG*) achieved significantly higher critique prediction accuracy and recommendation accuracy (Chen and Pu 2007c). Moreover, its interface displays multiple sample products along with a suggested critique. Motivated by these advantages, a new version of the hybrid critiquing system was designed that combines Pref-ORG and Example Critiquing (Chen and Pu 2007b).

Figure 10 shows the screenshots. Here we give an example to illustrate how the system models the user preferences and generates critiques. A user initially starts her search by specifying one or any number of preferences. Each preference is composed of one acceptable attribute value and its relative importance (weight). A preference structure is hence a set of (*acceptable attribute value, weight*) pairs of all participating main attributes. After a user specifies her initial preferences, the best matching product computed by the MAUT model will be returned at the top, and followed by the suggested critiques and sample products as produced by the preference-based



(**b**) The user-initiated example critiquing interface.

Fig. 10 Hybrid critiquing system (version 2): the combination of preference-based organization interface (Pref-ORG) and user-initiated critiquing facility (Chen and Pu 2007b, 2010)

organization algorithm (as described in Sect. 2.2). If the user is interested in one of the suggested critiques, she could click "Show All" to see more products under the critique. Among these products, the user can either choose one as her final choice, or

select a near-target and click "Better Features" to start a new round of critiquing. In the latter case, the user's preference model will be automatically refined to respect her current criteria.

On the other hand, if in the organization interface, no critique and product interest the user, she could switch to make the self-initiated critiquing by clicking on the button "Self specify criteria for 'Better Features'" (see Fig. 10). At this point, the Example Critiquing interface will be shown and facilitate the user to create her critiquing criteria. After she creates her own critiques, the system will also refine the user's preference model and return multiple tradeoff alternatives that best match to her self-specified critiquing criteria (as described in Sect. 2.3).

The user's action of selecting system-suggested critiques or making self-initiated critiquing completes one critiquing cycle, and it continues as long as the user wants to refine the results.

A comparative user study compared this new hybrid design with the first version. A total of 44 volunteers participated in the experiment (Chen and Pu 2007b). This user evaluation revealed that users more frequently picked suggested critiques in the second system, and significantly saved their task time and interaction effort. Moreover, both hybrid critiquing systems obtained high scores regarding participants' subjective decision confidence and purchase/return intentions, implying that users will likely possess positive attitude towards a system that comprises both system-suggested critiques and user-initiated critiquing support.

Above results hence verify the earlier hypothesis that hybrid critiquing systems can effectively address previous single systems' respective limitations, and allow users to reach high level of decision accuracy, while demanding low objective effort in making the decision. Furthermore, users subjectively prefer such kind of hybrid system, and intend to purchase the chosen product and even repeatedly use the system in the future.

4 Adoption of crituqing based recommenders in industry

So far we have described researchers' attempts to develop and test various critiquing based recommenders. A multitude of user studies also demonstrated that these systems effectively help users' decision process because they are able to adapt to users' changing needs and preferences. In this chapter, we describe the application of the critiquing method in two online websites. Since 2009, Amazon provides a new function, called "Fix this recommendation", for each item listed in the "recommendation for you" area. When a user clicks the link,¹ a pop-up window opens up and the user can indicate that "I own it", "not interested", or "don't use for recommendations". "Fix this recommendation" thus stimulates users to *critique* on the book that was suggested to them. The three subsequent critiquing actions allow users to tell the system how to improve the current recommendation: "I own it" means that the system should not ever suggest this book to me, "not interested" indicates that this book does not interest me, and "don't use for recommendations" alerts to the system that this book does not characterize my taste.

¹ "Fix this recommendation" is implemented as an embedded link on the interface.

Since 2010, the popular movie recommender Movielens also added an example critiquing function, called the "Movie Tuner" (Vig et al. 2011). With this function, the user is able to adjust the movies recommended to her by requesting more or less of particular qualities, for example "more action" or "less violent". After the critique is made, the system will recommend movies that meet the user's new requirements.

5 Future trends

Motivated by the historical review and survey of recent works on critiquing system development, we are able to suggest several directions for future investigation in this area.

5.1 Hybrid systems

Given the proven merits of different types of critiquing approaches and their combined effects in the hybrid systems, we believe that the hybrid approach could be further enhanced in the future. For instance, with recent advance in natural language processing and speech recognition technologies, the limitations of traditional natural dialog model could be potentially addressed, and its advantages (e.g., easy to use and simulating to the natural conversional style with a real salesperson) can be incorporated into existing hybrid systems, so as to further augment the system's enjoyability and user adoption degree.

5.2 Adaptive element

Another emerging trend will be further improving the systems' ability of adapting the critiquing process to the user's changing preferences. In this regard, some pioneer researches have started. Viappiani et al. (2007b) have been attempting to add the "adaptive suggestions" element in their Flat Finder, by which the system can learn from the user by observing her reaction to displayed examples. For instance, if the user is shown an option with value *subway* for attribute "transportation" and she has not stated any critique about transportation, the probability that the user has a preference for *subway* as transportation will decrease. In the next interaction cycle, the system will choose options with a different value. On the other hand, Chen (2009) has highlighted two typical preference conditions that a user will be likely involved during her entire decision process: "incomplete preferences" when the user has not stated any preferences on the major product features, and "conflicting preferences" when the user's stated preferences conflict with each other so that no product satisfies all of them. Therefore, the paper proposed an adaptive approach to selecting critiques, so as to explicitly tailor to the two preference conditions. Mahmood and Ricci (2009) have adopted the Reinforcement Learning (RL) technique to observe the responses of users in a conversational recommender, with the aim to maximize a numerical cumulative reward function that they use to model how much benefit the user is supposed to get from each recommendation session.

5.3 Critiquing systems for low-involvement product domains

As shown in Table 1, most existing critiquing-based recommender systems have been oriented to high-value (or called high-involvement) products such as PCs, digital cameras, and travel planning, since for these products consumers are more motivated to spend a significant amount of time in making critiques and rely on their own judgment in the selection process, with the purpose of avoiding any financial risks. However, few have studied whether the critiquing element could also be applicable in the lowinvolvement product domains such as music, movies, perfumes and wines, though users' risk avoidance and decision accuracy would not be demanded so high for these products. To fill in the gap, Pu et al. (2009) developed a new system-suggested critique generation tool, called Editorial Picked Critiques (EPC), with the assumption that users would spend less time in choosing low-risk products and are more likely to rely on public opinions or experts' advice to make decisions. The critiques were hence generated through the data mining technique to combine editorial opinions, popularity information and preference-based critiques. This method was tested in a perfume recommender, and demonstrated to significantly improve users' decision confidence. In the future, it will be interesting to follow this direction to further investigate whether similar benefits could bring to other kinds of low-risk goods.

6 Conclusion

This article provides a detailed review of the development of critiquing-based recommender systems in the past decade. Particularly, three branches of development were elaborated: systems with natural language dialogs, system-suggested critiquing systems where users select one of the proposed critiques as the feedback option, and userinitiated critiquing systems where users create critiques based on their own desires. Following the discussion of their respective pros and cons, we presented experimental results from comparing different approaches, and the derivation of hybrid critiquing systems that unified these approaches' advantages in a single platform. In particular, most of reported experiments were conducted to identify user experiences with these types of critiquing systems, which objective is indeed regarded more effective than purely testing the algorithms for judging the recommender's practical benefits (Konstan and Riedl 2012). In the last two Sects. 4 and 5, we described the recent adoption of critiquing recommend systems in online industries and pointed out three major emerging trends in this area: the incorporation of intelligent conversational elements, the improvement on systems' adaptation ability, and the application of critiquing systems in low involvement product domains.

References

Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Buneman, P., Jajodia, S. (eds.) Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data (SIGMOD'93), pp. 207–216. ACM, New York (1993)

- Burke, R.: Knowledge-based Recommender Systems. Encyclopedia of Library and Information Systems 69 (2000)
- Burke, R., Hammond, K., Cooper, E.: Knowledge-based navigation of complex information spaces. In: Proceedings of the 13th National Conference on Artificial Intelligence (AAAI'96), pp. 462–468 (1996)
- Burke, R., Hammond, K., Young, B.: The FindMe approach to assisted browsing. IEEE Expert: Intell. Syst. Their Appl. **12**, 32–40 (1997)
- Carenini, G., Poole, D.: Constructed preferences and value-focused thinking: implications for AI research on preference elicitation. AAAI-02 Workshop on Preferences in AI and CP: symbolic approaches, Edmonton (2000)
- Chen, L.: Adaptive tradeoff explanations in conversational recommenders. In: Proceedings of ACM Conference on Recommender Systems (RecSys'09), ACM, New York, pp. 225–228 (2009)
- Chen, L., Pu, P.: Evaluating critiquing-based recommender agents. In: Proceedings of Twenty-first National Conference on Artificial Intelligence (AAAI'06), Boston, pp. 157–162 (2006)
- Chen, L., Pu, P.: Hybrid critiquing-based recommender systems. In: Proceedings of International Conference on Intelligent User Interfaces (IUI'07), Hawaii, pp. 22–31 (2007a)
- Chen, L., Pu, P.: The evaluation of a hybrid critiquing system with preference-based recommendations organization. In: Proceedings of ACM Conference on Recommender Systems (RecSys'07), Minneapolis, Minnesota, pp. 169–172 (2007b)
- Chen, L., Pu, P.: Preference-based organization interfaces: aiding user critiques in recommender systems. In: Proceedings of International Conference on User Modeling (UM'07), Corfu, pp. 77–86 (2007c)
- Chen, L., Pu, P.: Interaction design guidelines on critiquing-based recommender systems. User Model. User-Adapt. Interact. J. (UMUAI) **19**(3), 167–206 (2009)
- Chen, L., Pu, P.: Experiments on the preference-based organization interface in recommender systems. ACM Trans. Comput.-Hum. Interact. (TOCHI) **17**(1), 1–33 (2010)
- Faltings, B., Pu, P., Torrens, M., Viappiani, P.: Designing example-critiquing interaction. In: Proceedings of International Conference on Intelligent User Interfaces (IUI'04), ACM, New York, pp. 22–29 (2004a)
- Faltings, B., Torrens, M., Pu, P.: Solution generation with qualitative models of preferences. Int. J. Comput. Intell. Appl. 20, 246–264 (2004b)
- Jurca, A.: Consumer-centered interfaces: customizing online travel planning. In: CHI '00 Extended Abstracts on Human Factors in Computing Systems, ACM, New York, pp. 93–94 (2000)
- Keeney, R., Raiffa, H.: Decisions with multiple objectives: preferences and value tradeoffs. Cambridge University Press, Cambridge (1976)
- Konstan, J.A., Riedl, J.: Recommender systems: from algorithms to user experience. User Model. User-Adapt. Interact. J. (UMUAI), 22 (2012)
- Linden, G., Hanks, S., Lesh, N.: Interactive assessment of user preference models: the automated travel assistant. In: Proceedings of International Conference on User Modeling (UM'97), pp. 67–78 (1997)
- Mahmood, T., Ricci, F.: Improving recommender systems with adaptive conversational strategies. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia (HT '09), ACM, New York, pp. 73–82 (2009)
- McCarthy, K., Reilly, J., McGinty, L., Smyth, B.: On the dynamic generation of compound critiques in conversational recommender systems. In: Proceedings of the Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'04), pp. 176–184 (2004a)
- McCarthy, K., Reilly, J., McGinty, L., Smyth, B.: Thinking positively: explanatory feedback for conversational recommender systems. In: Proceedings of the Workshop on Explanation in CBR at the Seventh European Conference on Case-Based Reasoning, Madrid, pp. 115–124 (2004b)
- McCarthy, K., McGinty, L., Smyth, B., Reilly, J.: A live-user evaluation of incremental dynamic critiquing. In: Proceedings of International Conference on Case-based Reasoning (ICCBR'05), pp. 339–352 (2005a)
- McCarthy, K., McGinty, L., Smyth, B., Reilly, J.: On the evaluation of dynamic critiquing: a large-scale user study. In: Proceedings of the Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference, Pittsburgh, pp. 535–540 (2005b)
- McCarthy, K., Reilly, J., McGinty, L., Smyth, B.: Experiments in dynamic critiquing. In: Proceedings of International Conference on Intelligent User Interfaces (IUI'05), San Diego, pp. 175–182 (2005c)
- Payne, J.W., Bettman, J.R., Johnson, E.J.: The adaptive decision maker. Cambridge University Press, Cambridge (1993)

- Payne, J.W., Bettman, J.R., Schkade, D.A.: Measuring constructed preference: towards a building code. J. Risk Uncertain. 19(1–3), 243–270 (1999)
- Pu, P., Chen, L.: Integrating tradeoff support in product search tools for e-commerce sites. In: Proceeding of the ACM Conference on Electronic Commerce (EC'05), Vancouver, pp. 269–278 (2005)
- Pu, P., Chen, L.: Trust building with explanation interfaces. In: Proceedings of International Conference on Intelligent User Interfaces (IUI'06), Sydney, pp. 93–100 (2006)
- Pu, P., Chen, L., Kumar, P.: Evaluating product search and recommender systems for e-commerce environments. Elec. Commer. Res. J. 8(1–2), 1–27 (2008)
- Pu, P., Faltings, B.: Enriching buyers' experiences: the SmartClient approach. In: Proceedings of the SIG-CHI Conference on Human Factors in Computing Systems (CHI'00), ACM, New York, pp. 289–296 (2000)
- Pu, P., Faltings, B.: Personalized navigation of heterogeneous product spaces using SmartClient. In: Proceedings of the International Conference on Intelligent User Interfaces (IUI'02), pp. 212–213 (2002)
- Pu, P., Faltings, B.: Decision tradeoff using example critiquing and constraint programming. Special Issue on User-Interaction in Constraint Satisfaction, CONSTRAINT 9(4), 289–310 (2004)
- Pu, P., Faltings, B., Chen, L., Zhang, J., Viappiani, P.: Usability guidelines for product recommenders based on example critiquing research.In: Recommender systems handbook, ISBN: 978-0-387-85819-7, pp. 511–546 (2011)
- Pu, P., Kumar, P.: Evaluating example-based search tools. In: Proceeding of the ACM Conference on Electronic Commerce (EC'04), New York, pp. 208–217 (2004)
- Pu, P., Zhou, M., Castagnos, S.: Critiquing recommenders for public taste products. In: Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys'09), New York, pp. 249–252 (2009)
- Reilly, J., McCarthy, K., McGinty, L., Smyth, B.: Dynamic critiquing. In: Proceedings of European Conference on Case-based Reasoning (ECCBR'04), Madrid, pp. 763–777 (2004)
- Reilly, J., McCarthy, K., McGinty, L., Smyth, B.: Explaining compound critiques. Artif. intell. Rev. 24(2) (2005a)
- Reilly, J., McCarthy, K., McGinty, L., Smyth, B.: Incremental critiquing. J. Knowl.-Based Syst. 18(4–5), 143–151 (2005)
- Reilly, J., Zhang, J., McGinty, L., Pu, P., Smyth, B.: Evaluating compound critiquing recommenders: a real-user study. In: Proceedings of ACM Conference on Electronic Commerce (EC'07), San Diego, pp. 114–123 (2007)
- Shearin, S., Lieberman, H.: Intelligent profiling by example. In: Proceedings of Conference on Intelligent User Interfaces (IUI'01), Santa Fe, pp. 145–151 (2001)
- Shimazu, H.: ExpertClerk: navigating shoppers' buying process with the combination of asking and proposing. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI'01), Seattle (2001)
- Smyth, B., McGinty, L.: An analysis of feedback strategies in conversational recommenders. In: the Fourteenth Irish Artificial Intelligence and Cognitive Science Conference, Dublin (2003)
- Thompson, C.A., Goker, M.H., Langley, P.: A personalized system for conversational recommendations. J. Artif. Intell. Res. 21, 393–428 (2004)
- Torrens, M., Faltings, B., Pu, P.: SmartClients: constraint satisfaction as a paradigm for scaleable intelligent information systems. Int J Constraints 7(1), 49–69 (2002)
- Torrens, M., Weigel, R., Faltings, B.: Java constraint library: bringing constraints technology on the Internet using the Java language. In: Workshop of National Conference on Artificial Intelligence (AAAI), pp. 10–15 (1997)
- Tversky, A., Simonson, I.: Context-dependent preferences. Manag. Sci. 39(10), 1179–1189 (1993)
- Viappiani, P., Faltings, B., Pu, P.: Evaluating preference-based search tools: a tale of two approaches. In: Proceedings of the Twenty-first National Conference on Artificial Intelligence (AAAI'06), Boston, pp. 205–211 (2006)
- Viappiani, P., Faltings, B., Pu, P.: Preference-based search using example-critiquing with suggestions. J. Artif. Intell. Res. 27, pp. 465–503 (2007a)
- Viappiani, P., Pu, P., Faltings, B.: Conversational recommenders with adaptive suggestions. In: Proceedings of the 2007 ACM Conference on Recommender Systems (RecSys '07), Minneapolis, pp. 89–96 (2007b)
- Vig, J., Sen, S., Riedl, J.: Navigating the tag genome. In: Proceedings of the 16th International Conference on Intelligent User Interfaces (IUI'11), Palo Alto, pp. 93–102 (2011)

- Zhang, J., Pu, P.: A comparative study of compound critique generation in conversational recommender systems. In: Proceedings of International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'06), Dublin, pp. 234–243 (2006)
- Zhang, J., Jones, N., Pu, P.: A visual interface for critiquing-based recommender systems. In: Proceedings of ACM Conference on Electronic Commerce (EC'08), Chicago, pp. 230–239 (2008)

Author Biographies

Li Chen is Assistant Professor of Computer Science at Hong Kong Baptist University. She obtained her PhD degree in Human Computer Interaction at Swiss Federal Institute of Technology in Lausanne (EPFL), and Bachelor and Master Degrees in Computer Science at Peking University, China. Her research interests are mainly in the areas of human-computer interaction, user-centered development of recommender systems and e-commerce decision supports. Her co-authored papers have been published in journals and conferences on e-commerce, artificial intelligence, intelligent user interfaces, user modeling, and recommender systems. She is the vice chair of the ACM Hong Kong Chapter.

Pearl Pu obtained her Master and PhD degrees from the University of Pennsylvania in artificial intelligence and computer graphics. She currently leads the HCI Group in the School of Computer and Communication Sciences at the Swiss Federal Institute of Technology in Lausanne (EPFL). Her research is multi-disciplinary and focuses on issues in the intersection of human computer interaction, artificial intelligence, and behavioral decision theories. Her work was concerned with decision support systems, electronic commerce, online consumer decision behavior, product recommender systems, content-based product search, travel planning tools, trust-inspiring interfaces for recommender agent, music recommenders, and user technology adoption. She is associate editor for the IEEE Transactions on Multimedia and editorial board member of User Modeling and User-Adapted Interaction.