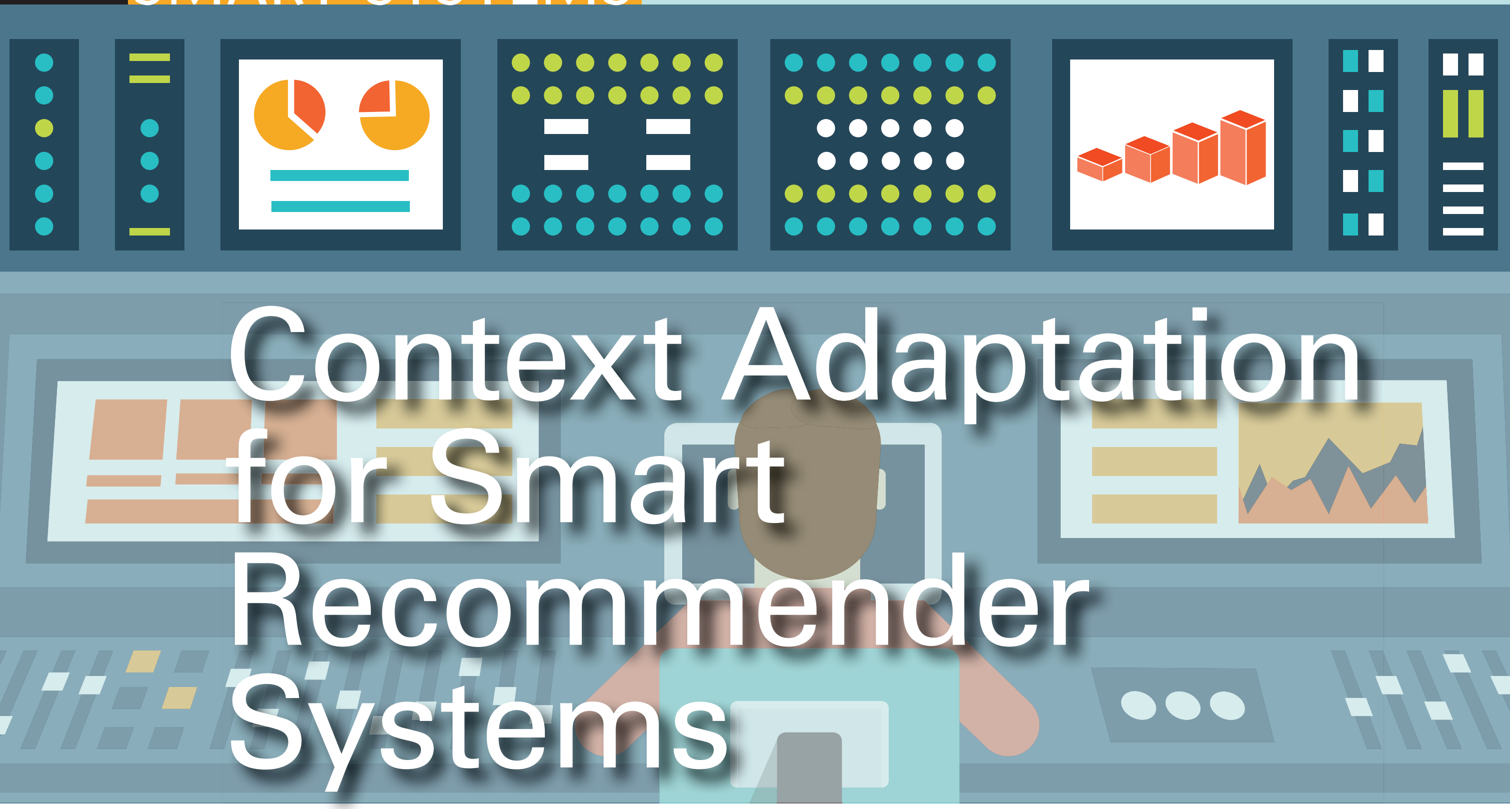

Context Adaptation for Smart Recommender Systems

Fanjuan Shi, Chirine Ghedira, and Jean-Luc Marini, *Université Jean Moulin Lyon 3, France*

Contextual factors are an important mediator for improving recommender system performance. The authors' real-time, context-aware recommendation model can identify users' state of mind and budget based on clickstream data to improve recommendations for e-commerce.

E-commerce recommender systems (RSs) are widely deployed by online merchants[1] to facilitate the consumer shopping process.[2] Many RSs today can provide personalized recommendations based on users' historical behavior.[3] A clicked recommendation can increase website sales and enhance customer satisfaction. Hence, improving RS performance is considered an important topic for researchers and online merchants. Most existing techniques focus on developing and optimizing algorithms,[4] such as collaborative filtering algorithms, content-based algorithms, and hybrid systems.[5,6] Other researchers improve RS performance by enhancing user experience, and several proposals look at user interface enhancement[7] and information presentation.[8–10]

Despite various efforts to improve RS performance, recommendation click-through rates (CTRs) and purchase rates (PRs), which measure how many proposed recommendations end up being clicked (purchased), are low (see http://tinyurl.com/psjuv5v). Researchers are finding that RS performance is associated not only with algorithms and the user interface but also with contextual factors that mediate users' decisions.[11]

To better understand how contextual factors apply to e-commerce, let's take the following motivation case, which illustrates the contextual factors that could be essential in e-commerce recommender systems, and why. Consider Tom and Joe, two users who visit an e-commerce website. Tom plans to spend 600 euros to buy a bike this week, whereas Joe wants to buy a birthday

Published by the IEEE Computer Society

present for his best friend without having a clear idea of what. By chance, they type the same keyword (mountain bike) in the search engine and browse the same products. The RS on this website is equipped with a collaborative filtering algorithm that proposes new items based on items previously viewed by users. As a result, the RS recommends a couple of bikes to Tom and Joe. Tom is interested in one of these bikes and starts to browse it immediately. Conversely, Joe quickly bounces to the video game category, because he thinks that a PlayStation game might be a better gift. Besides, bikes seem too expensive.

Judging from his behavior, Joe seems interested in bikes, but in reality, he's trying to find a suitable birthday present at a reasonable price. If the RS were aware of this, it could have proposed different kinds of gifts for Joe to consider. In that case, Joe might have clicked or even bought a recommended item.

The motivation case indicates two challenges for e-commerce RSs. First, despite having the same search and browse behavior, users might have different considerations. Recommendations adapted to these considerations might be more interesting for users. Second, a user's budget is an important constraint for shopping. RS performance might improve if recommendations take budget into account.

To address these challenges, we propose an original, real-time, context-aware e-commerce RS whose performance can be constantly improved via a self-learning mechanism. Our work proposes new contextual factors for e-commerce RSs, presents a novel method to track users' real-time shopping context, and provides an autonomous mechanism to enhance RS performance.

Enhancing RSs with Context

To see how RSs can be further enhanced by context, we must follow two tracks of study—namely, defining and tracking e-commerce-specific contextual factors, and finding a way to measure RS performance improvements.

Context for E-Commerce RSs

Context is defined as any information that can be used to characterize the situation of an entity (person, place, or object) that's considered relevant to the interaction between users and an application.[12] Such information can be about the circumstances, objects, or conditions that surround a user.[6,13] Many researchers have attempted to classify context. Bill Schilit, Norman Adams, and Roy Want proposed three categories of context[14]: computing (such as cost), user (such as social situations), and physical (such as temperature). Gediminas Adomavicius and his colleagues classified context by data availability and stability.[13] Depending on the application domain, contextual factors can be different.[6,13] Although interesting and important to RSs, these general contextual factors aren't sufficient. In fact, we focus on identifying real-time, e-commerce-specific contextual factors.

In our research, we define context as the mental condition and physical constraints that affect users' shopping decisions. Based on this definition, we propose two new contextual factors: users' real-time state of mind (RSOM) and their current budget.

RSOM. State of mind is defined as a person's mood or mental state at a particular time (see www.collinsdictionary.com/dictionary/english/state-of-mind). RSOM detection is a new topic. Some researchers have studied the relationship between RSOM and RS usage. For instance, Gerhard Fischer found that RS usage is associated with the time when recommendations are made[15]; Shuk Ho and Kar-Yan Tam proposed that "time to recommend" should be determined at different stages of decision making.[16]

These approaches detect RSOM based on semantic and nonsemantic approaches. Indicators such as Minkowski distance, cosine similarity, the Pearson correlation coefficient, and Jaccard similarity are used in their analyses. However, most of these solutions have several limitations. First, the results of such semantic approaches as page similarity analysis are sensitive to the quality of the product taxonomy, thus limiting their predictive accuracy. Second, for an e-commerce website with a broad assortment of goods, a semantic approach might lose predictive power due to data sparsity and the high-dimensionality problem. Third, nonsemantic approaches such as path similarity analysis don't consider the path's hierarchical structure. As a result, they can't predict if users need more general or specific information. Finally, these approaches don't consider the frequent changes to product catalogs and taxonomies in e-commerce websites. This limitation can be an issue in real e-commerce practice. Hence, a new RS to address these issues is essential.

Users' current budget. Although many e-commerce websites allow users to filter search results by price, budget hasn't been comprehensively discussed in e-commerce RS research, nor do we find RSs that actively adapt recommendations to users' budgets. Given that both the definition of context[14] and consumer theory[17] suggest that budget is a key decision criterion in making purchases, it will be interesting to explore how it affects e-commerce RS usage.

RS Evaluation Criteria and Metrics

As for evaluation criteria and metrics, although most researchers evaluate RSs based on user feedback and actions,[17] there are still some differences between academic research and industrial studies.

Indeed, due to infrastructural, financial, and data availability constraints, academic researchers prefer to conduct experiments with a group of selected users in the lab. Although such methods are effective and necessary, it's inevitable that participating users each have their own way of perceiving and interpreting the evaluation criteria. Sometimes, such perceptual deviation can affect the credibility and reliability of the experiment.[18,19] From an industrial viewpoint, this method isn't sufficient if we want to take RSs from the lab to the market.[20]

Many researchers base their research on clickstream data, a digital record of users' online behaviors.[21] Behavioral indicators such as time, browsing path, users' choice, mouse actions, and keystrokes are collected unobtrusively and reflect users' intent and decision-making processes.[22] Metrics such as CTR, PR, frequency, and magnitude of browsing behavior are used to measure RS performance. We thus developed our RS based on clickstream data analysis.

Toward a Context-Adapted Intelligent RS

Our research is based on a French e-commerce website (FECW) that sells a broad assortment of products, such as sports equipment, consumer electronics, white goods, and cultural products. FECW is equipped with a hybrid RS named MARS, which presents recommendations simultaneously to users based on two algorithms: Delta and Sigma. Delta is a content-based algorithm that assumes that users give similar ratings to items with similar characteristics. So, if users browse a bike, Delta will propose some bikes. Items recommended by Delta are always directly related to users' current search interests. When users are searching for specific items, Delta can be a good decision-support tool.

Sigma is a collaborative algorithm. It assumes that users who have similar tastes will rate things similarly. For example, if user$_x$ likes items A, B, and C and user$_y$ likes items A and C, Sigma is likely to recommend B to user$_y$, given that the two users have the same tastes. Items recommended by Sigma can be different from users' current search interests. For example, if users browse a bike, Sigma might recommend some bike accessories. When users want to extend their browsing scope, Sigma can be a good assistant. To summarize, the MARS recommendation methods and results are complementary, but aren't adapted to users' state of mind or budget.

Therefore, we propose a context-aware version of MARS (CAMARS). Briefly, CAMARS adds two new modules: an *RSOM detector* (RSOMD) and a *users' budget estimator* (UBE). RSOMD monitors users' state of mind in real time and chooses the most relevant algorithm accordingly; UBE estimates users' real-time budget, which will make recommendations more relevant. Figure 1 illustrates the differences between the two RS algorithms.

Before introducing the new modules, let's look at how we create a category identification (CID). To overcome the shortcomings of page- and pattern-similarity approaches, we code category pages based on a website's hierarchical structure. First, all product categories, subcategories, and items are aligned so that they are hierarchically comparable. Then, let $CIDk = \sum_1^L 10^{2(L-1)} * (10 + \beta_L)$, where L is the hierarchical level of category k (1 denotes the lowest level), and β_L is the serial number of category k in its parent category, $\beta_L \in [1, 89]$. Finally, we define the CID of an item page as the CID of its direct parent category. As a result, a CID reflects the hierarchical path of a category (or item). For example, for a mountain bike item (CID: 18602347) and a mudguard (CID:

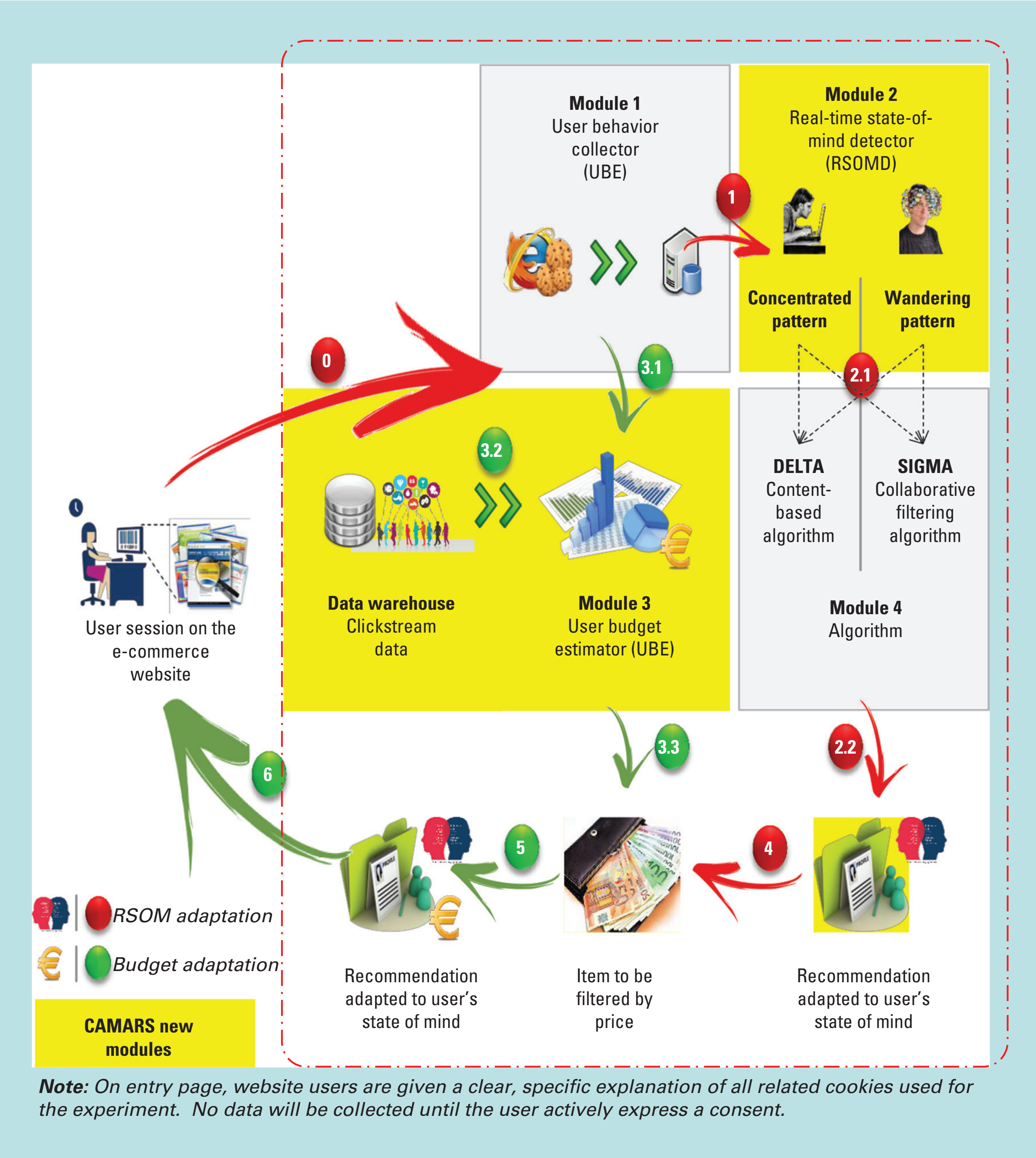

Figure 1. CAMARS architecture. The real-time state-of-mind detector (RSOMD) and users' budget estimator (UBE) make the recommender system context aware and the recommendations more relevant.

18607522), 18(sports) represents a common parent category, 60(cycling products) represents its child category, 23(bikes) and 75(accessories) represent subcategories, and 47(mountain bikes) and 22(mudguard) represent segments. Accordingly, user$_x$'s browsing path can be defined as a series of CIDs arranged in chronological order: $C_x = \{CID_x^1, CID_x^2, \ldots CID_x^n\}$.

This approach offers several benefits. First, it enables RSs to compare several CIDs at a given time to identify the common feature. The result is less sensitive to product taxonomy, without data sparsity or high-dimensionality issues. Second, because it keeps a category's hierarchical information, it can predict whether users are looking for specific or general

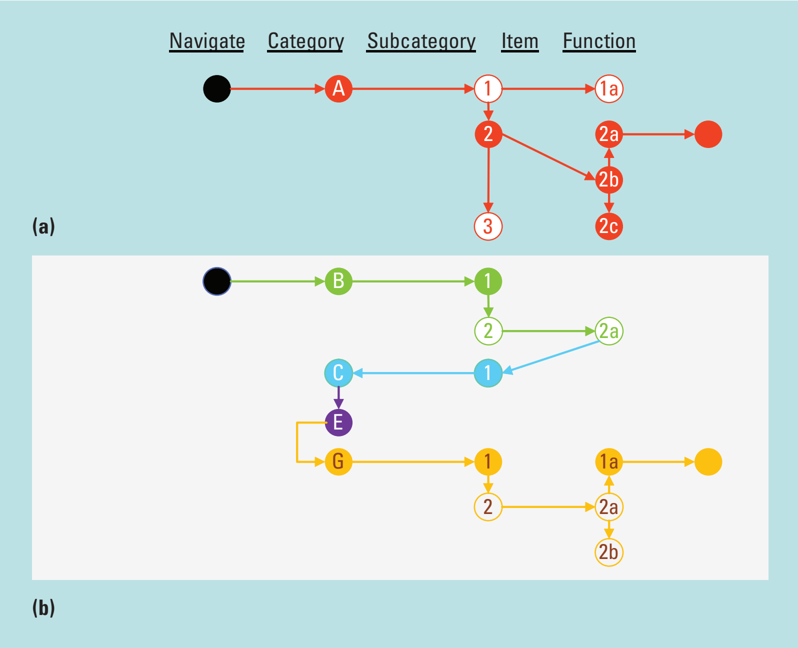

Figure 2. User browsing patterns. In (a) the concentrated pattern, users have a specific search focus. In (b) the wandering pattern, users don't demonstrate a specific interest.

information. Finally, this coding method is scalable; thanks to the way it's coded, theoretically it can cater to any broad assortment without repetition.

RSOMD

Based on the page-coding approach, RSOMD calculates the standard error (SE) of a given series of CIDs. Because a CID contains hierarchical information, its SE can be used to determine whether users are looking for general (high SE value) or specific (low SE value) information. Moreover, if the SE is decreasing, the system can infer that users are starting to focus on a particular topic, which can be identified by the last few CIDs in the series.

Using historical clickstream data from 238,874 customers (602,861 sessions) who used MARS from May to October 2013, we explored the relationship between the SE of CIDs and the recommendation CTR. We can derive two major findings.

First, 52 percent of the sessions recorded an SE lower than 100, and 39 percent recorded an SE higher than 10,000. This indicates that there are two main browsing patterns: concentrated and wandering. In the concentrated pattern, users have a specific search focus. As a result, the CIDs of their webpages are similar, and the SE of the CIDs is low. This indicates that users have a concentrated RSOM. In the wandering pattern, users don't demonstrate a specific interest. Accordingly, they bounce between different categories and produce a high SE value. This indicates that users have a wandering RSOM. Figure 2 illustrates examples of these two patterns.

Second, RS performance is associated with users' RSOM. In a concentrated pattern, CTR_{Delta} is 0.76 percent, and CTR_{Sigma} is only 0.17 percent. However, in a wandering pattern, CTR_{Delta} is 0.19 percent, and CTR_{Sigma} is nearly 0.80 percent. This suggests that if the RS can automatically choose the relevant algorithm based on users' RSOM, its performance can be potentially enhanced.

Based on these findings, we propose a self-learning mechanism to select algorithms based on users' RSOM. At the beginning, each algorithm is given the same probability of being chosen for a given RSOM. During this "learning phase," recommendations made by certain algorithms are clicked more often than others in a given RSOM, suggesting that these algorithms are more relevant to this RSOM. After the learning phase, the RS will take such differences into consideration. As a result, more relevant algorithms will have a higher probability of being chosen.

UBE

Before determining users' budget, a key concept must be defined: users' recent interests, or *most recent common category* (MRCC). A common category (CC) is the common fragment of several CIDs, starting from the first digit to the last common digit. The recent common category (RCC) can be obtained by comparing the last CID with the previous CID. If several RCCs are identified, the one with more common fragments is considered the MRCC. For example, let C = {ABCD, ABGE, ABCF, ABGH, ABCK, ABGL}, then CC = {A###, AB##, ABC#, ABG#}, RCC = {A###, AB##, ABG#}, and MRCC = {ABG#}.

A user's budget is defined as a price range, which is determined by the average price of all items in the user's MRCC ($Price_{\text{MRCC}}$) and a coefficient θ:

$$\begin{cases} Budget_{\text{user}} \in \left[\theta_{\text{lower}} * Price_{\text{MRCC}},\ \ \theta_{\text{upper}} * Price_{\text{MRCC}}\right] \\ Price_{\text{MRCC}} = \dfrac{\sum_{j=1}^{n} price_j}{n},\ Price_j - \text{price of item } j \text{ in MRCC} \end{cases}$$

$Price_{\text{MRCC}}$ can be calculated based on the real-time clickstream data; θ is a budget flexibility coefficient to be estimated using the historical clickstream data of MARS users.

We let $\phi = Price_{PR} / Price^{*}_{\text{MRCC}}$, where $Price_{\text{PR}}$ is the price of a recommended item purchased by users, and $Price^{*}_{\text{MRCC}}$ is the average price of items in MRCC before they are purchased. Then, we examine the relationship between ϕ and the purchase frequency/rate by plotting the data into Figure 3. We conduct a similar exercise to identify the relationship between ϕ and the click frequency/rate.

More than 60 percent of recommendation clicks and 70 percent of recommendation purchases are gathered in an interval where ϕ = [0.8, 1.2]. These findings suggest that budget plays an important role in users' item choice. In our case study, the initial θ is set as [0.8, 1.2] because this interval records the highest CTR/PR. Then, UBE runs this process periodically based on historical clickstream data, so that recommendations made by the RS can be more relevant to users' needs. Depending on the privacy policy, the budget flexibility coefficient θ can be estimated at the personal or group level.

Experimental Settings and Empirical Results

To validate our method, we conducted a live experiment at FECW. The idea was to measure the performance of the MARS versus CAMARS RSs under the same conditions and during the same time (that is, an A/B test) to see if contextual factors can enhance system performance. Accordingly, users were randomly allocated to one of the following groups when they visited FECW (p = 25 percent). In Group 1, MARS was used as the baseline system for benchmarking. In Group 2, CAMARS made recommendations based on the user's browsing patterns and MRCC (shopping intent). In Group 3, CAMARS estimated the user's budget and adapted recommendations to it. In Group 4, both mechanisms were activated.

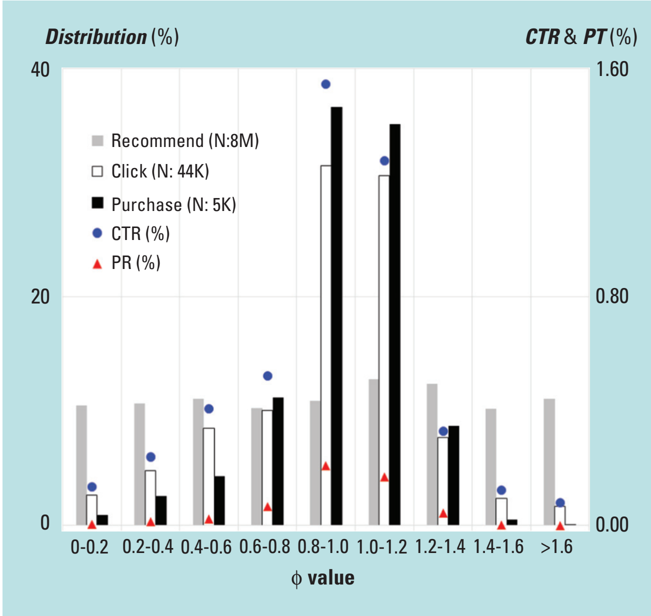

Figure 3. Distribution of displayed, clicked, and purchased recommendations (left axis) and the click-through rate (CTR) and purchase rate (PR) (right axis) by ϕ value. Recommended items were more often clicked and purchased when their price was within users' budget range.

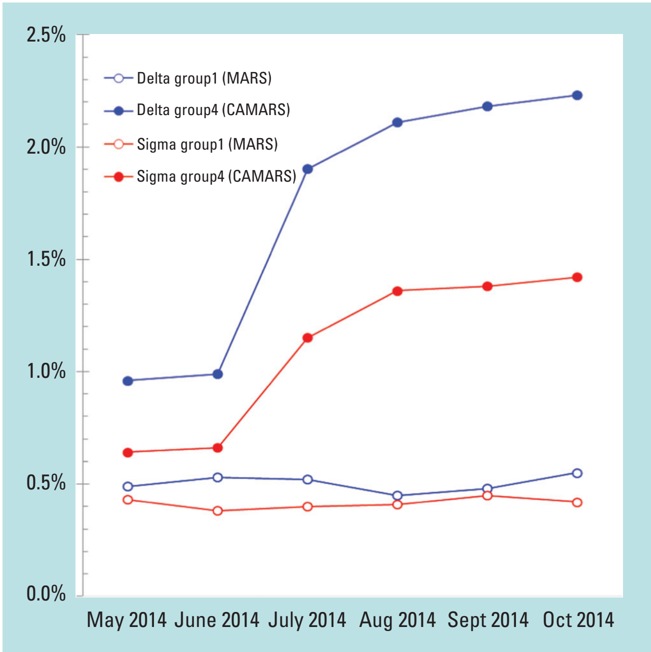

Figure 4. Empirical results. Compared to the stable performance of MARS, significant and constant improvements can be observed in the CAMARS groups.

From 1 May to 31 October 2014, more than 253,000 users participated in the experiment. Table 1 and Figure 4 illustrate the key performance indicator (KPI) for both RSs.

Table 1. Key performance indicators (KPIs) for MARS and CAMARS systems.

KPI: Delta	Group 1: MARS	Group 2: CAMARS (RSOM)	Group 3: CAMARS (Budget)	Group 4: CAMARS (Both)
Probability	25%	25%	25%	25%
No. of sessions	160,838	160,838	160,838	160,838
No. of recommendations	2,359,403	1,560,539	2,370,992	1,538,471
No. of click-throughs	12,186	10,072	23,505	26,805
Click-through rate (CTR)	0.52%	0.65%	0.99%	1.74%
No. of purchases	1,332	971	3,024	2,826
Purchase rate (PR)	0.06%	0.06%	0.13%	0.18%
Page view time (minutes)	1.3	1.7	4.2	4.6
Mouse click and scroll (no. of times)	6.4	8.2	21.5	23.7
Mouse movement (pixels)	10,937	15,743	37,512	41,366
KPI: Sigma				
Probability	25%	25%	25%	25%
No. of sessions	160,838	160,838	160,838	160,838
No. of recommendations	2,359,403	1,360,385	2,370,992	1,339,958
No. of click-throughs	9,825	8,650	16,094	14,668
CTR	0.42%	0.64%	0.68%	1.09%
No. of purchases	737	389	1,247	913
PR	0.03%	0.03%	0.05%	0.07%
Page view time (minutes)	1.7	3.1	5.3	5.4
Mouse click and scroll (no. of times)	8.4	20.4	30.9	28.5
Mouse movement (pixels)	5,339	12,691	6,994	25,311

We can draw several conclusions based on the experiment's result. First, the enhancements to the number of click-throughs, number of purchases, CTR, and PR (see Table 1) confirm that integrating users' state of mind and budget into the recommendation can significantly enhance RS usage. Second, behavioral indicators such as page view time, mouse actions, and mouse movements demonstrate that users are more active when they browse items recommended by CAMARS. This suggests that users are more interested in the context-adapted recommendation. Finally, monthly CTR (Figure 4) demonstrates that the CTR of CAMARS continues to improve, whereas the CTR of MARS fluctuates. This indicates that the self-learning mechanism is effective, and CAMARS is constantly learning from its previous success to improve its predictive power.

Our findings have several implications for online merchants and marketers. First, a hybrid RS taking into account different kinds of Web browsing patterns might be better. As our experiment indicates, one algorithm can't meet all needs at the same time. Deploying complementary algorithms in RSs could help fulfill diversified user needs and enhance user satisfaction.

Second, different contexts require RSs to play different roles. In a concentrated pattern, users have a clear need, and the role of the RS is to help them find what they want. In a wandering pattern, users don't have a specific need, and they browse different webpages to look for ideas. In such circumstances, recommending novel or popular items might be a good way to enhance user satisfaction and stickiness.

Third, contextual factors can be used as marketing tools. One example is upselling. To increase customer basket value, many online merchants provide incentive policies when consumer spending reaches a certain amount (for example, free delivery or a cash coupon for their next visit). Sometimes, users see a little gap between their spending and the threshold. By viewing this gap as part of the user's budget, RSs can make recommendations to enhance upselling opportunities and customer satisfaction. The other example is cross-selling. When users are searching for specific items (concentrated pattern), recommending complementary products might not attract their attention. However, recommending accessories and complementary products can be really interesting when users have put an item into their shopping cart, because they are more relaxed and hence more open to suggestions at that moment.

Regarding system advancement, our future research will focus on three aspects: enhancing the RSOMD module to cater to more algorithms and more sophisticated shopping contexts; provisioning a scalable solution for websites that must respond to a big volume of user requests simultaneously; and extending the system to a multiscreen environment to address challenges of personalized recommendation when users bounce between different devices and channels.

Acknowledgments

We thank our industrial partner for their involvement in this research project. We also thank the reviewers for their valuable remarks and contributions to making the article suitable for IT Professional.

References

1. R. Burke, "Hybrid Web Recommender Systems," *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, eds., vol. 4321, Springer, 2007, pp. 377–408.
2. H.F. Wang and C.T. Wu, "A Strategy-Oriented Operation Module for Recommender Systems in E-Commerce," *Computers & Operations Research*, vol. 39, no. 8, 2012, pp. 1837–1849.
3. K.N. Wei, J.H. Huang, and S.H. Fu, "A Survey of E-Commerce Recommender Systems," *Proc. 2007 Int'l Conf. Service Systems and Service Management*, 2007, pp. 1–5.
4. B.P. Knijnenburg et al., "Explaining the User Experience of Recommender Systems," *User Modeling and User-Adapted Interaction*, vol. 22, nos. 4–5, 2012, pp. 441–504.
5. J.B. Schafer, J. Konstan, and J. Riedl, "Recommender Systems in E-Commerce," *Proc. 1st ACM Conf. Electronic Commerce*, 1999, pp. 158–166.
6. J. Boehmer, Y. Jung, and R. Wash, "E-Commerce Recommender Systems," *Int'l Encyclopedia of Digital Communication and Society*, Wiley, 2015.
7. P. Pu, L. Chen, and R. Hu, "A User-Centric Evaluation Framework for Recommender Systems," *Proc. 5th ACM Conf. Recommender Systems*, 2011, pp. 157–164.
8. L. Chen and P. Pu, "Experiments on User Experiences with Recommender Interfaces," *Behaviour & Information Technology*, vol. 33, no. 4, 2014, pp. 372–394.
9. R. Hu and P. Pu, "Enhancing Recommendation Diversity with Organization Interfaces," *Proc. 16th Int'l Conf. Intelligent User Interfaces*, 2011, pp. 347–350.
10. B. Scheibehenne, R. Greifeneder, and P.M. Todd, "Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload," *J. Consumer Research*, vol. 37, no. 3, 2010, pp. 409–425.
11. G. Adomavicius et al., "Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach," *ACM Trans. Information Systems*, vol. 23, no. 1, 2005, pp. 103–145.
12. A.K. Dey, G.D. Abowd, and D. Salber, "A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications," *Human-Computer Interaction J.*, special issue on context-aware computing, vol. 16, no. 2, 2001, pp. 97–166.
13. G. Adomavicius et al., "Context-Aware Recommender Systems," *Recommender Systems Handbook*, Springer, 2011, pp. 217–253.
14. B. Schilit, N. Adams, and R. Want, "Context-Aware Computing Applications," *Proc. 1st Workshop Mobile Computing Systems and Applications*, 1994, pp. 85–90.
15. G. Fischer, "Context-Aware Systems: The 'Right' Information, at the 'Right' Time, in the 'Right' Place, in the 'Right' Way, to the 'Right' Person," *Proc. Int'l Working Conf. Advanced Visual Interfaces*, 2012, pp. 287–294.
16. S.Y. Ho and K.Y. Tam, "An Empirical Examination of the Effects of Web Personalization at Different Stages of Decision Making," *Int'l J. Human-Computer Interaction*, vol. 19, no. 1, 2005, pp. 95–112.
17. A. Deaton and J. Muellbauer, "Consumer Demand Analysis—The Limits to Choice," *Economics and Consumer Behavior*, 1st ed., Cambridge Univ. Press, 1980, pp. 14–24.
18. J.L. Herlocker et al., "Evaluating Collaborative Filtering Recommender Systems," *ACM Trans. Information Systems*, vol. 22, no. 1, 2004, pp. 5–53.
19. G. Shani and A. Gunawardana, "Evaluating Recommendation Systems," *Recommender Systems Handbook*, Springer, 2011, pp. 257–297.
20. J. Picault et al., "How to Get the Recommender Out of the Lab?" *Recommender Systems Handbook*, Springer, 2011, pp. 333–365.

21. A.L. Montgomery et al., “Modeling Online Browsing and Path Analysis Using Clickstream Data,” *Marketing Science*, vol. 23, no. 4, 2004, pp. 579–595.
22. R.E. Bucklin and C. Sismeiro, “Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing,” *J. Interactive Marketing*, vol. 23, no. 1, 2009, pp. 35–48.

Fanjuan Shi *is a PhD candidate in the Magellan laboratory at the Institut d’Administration des Entreprises of Lyon, Université Jean Moulin Lyon 3, France. His interests include online consumer modeling and profiling, user-centric systems, context-aware recommender systems, and cross-channel shopping behavior. Contact him at fanjuan.shi@etu.univ-lyon3.fr.*

Chirine Ghedira *is a full professor of computer science and head of the Information Systems research group in the Magellan laboratory at Institut d’Administration des Entreprises of Lyon, Université Jean Moulin Lyon 3, France. Her interests include service-oriented architectures and computing; interoperability; complex, autonomic, and adaptive systems; and context-aware computing, along with data services, privacy, and cloud computing. Contact her at chirine.ghedira-guegan@univ-lyon3.fr.*

Jean-Luc Marini *is an adjunct professor of computer science in the Magellan laboratory at Institut d’Administration des Entreprises of Lyon, Université Jean Moulin Lyon 3, France. His research interests include artificial intelligence, personalized recommender systems, and search engines. Contact him at jean-luc.marini@univ-lyon3.fr.*