Consumer Preference Elicitation of Complex Products using Fuzzy Support Vector Machine Active Learning

Dongling Huang and Lan Luo

Web Appendix A: Pseudo Code of the Algorithm

For consumer $j = 1: J$

Step 1:

if $j = 1$

Initialize by $\tilde{\mathbf{w}}_j^0$ using

$$\tilde{\mathbf{w}}_j^0 = \tilde{\mathbf{w}}_j^I = \underset{\mathbf{w}_j^I, \alpha_j^I, \xi_j^i}{\arg\min} \quad \left[\frac{1}{2}\left\|\mathbf{w}_j^I\right\| + C\sum_{i=1}^{I}\xi_j^i\right] s.t. \quad y_j^i\left(\mathbf{w}_j^I * \mathbf{x}_j^i - \alpha_j^I\right) \geq 1 - \xi_j^i, \quad \xi_j^i \geq 0, \quad I = 2, \quad i = 1,2$$

else

Initialize $\tilde{\mathbf{w}}_j^0$ by using $\tilde{\mathbf{w}}_j^0 = \frac{1}{j-1}\sum_{j'=1}^{j-1} s(j, j') * \tilde{\mathbf{w}}_{j'}$ where $s(j, j') = \frac{\mathbf{c}_j^{'} * \mathbf{c}_{j'}}{\left\|\mathbf{c}_j\right\| * \left\|\mathbf{c}_{j'}\right\|}$.

Step 2:

Identify the must have or unacceptable features, and form the candidate profile pool N_j.

Step 3:

For $q1 = 1:Q1$,

Step 3.1:

u_j^i is obtained using
$$u_j^i = \begin{cases} 1 - \dfrac{\left\|\mathbf{x}_j^i - \overline{\mathbf{x}}_{j+}^I\right\|}{r_{j+}^I + \eta} & if\ y_j^i = 1 \\ 1 - \dfrac{\left\|\mathbf{x}_j^i - \overline{\mathbf{x}}_{j-}^I\right\|}{r_{j-}^I + \eta} & if\ y_j^i = -1 \end{cases} \qquad i = 1,2,\ldots,I$$
, where

$r_{j\pm}^I = \max_{i \in N_{j\pm};}(\left\|\mathbf{x}_j^i - \overline{\mathbf{x}}_{j\pm}^I\right\|)$, $\overline{\mathbf{x}}_{j\pm}^I = \frac{1}{N_{j\pm}^I}\sum_{i \in N_{j\pm}} \mathbf{x}_j^i$, $N_{j+}^I = \{i \mid y_j^i = 1\}$ and $N_{j-}^I = \{i \mid y_j^i = -1\}$

Step 3.2:

Calculate $\tilde{\mathbf{w}}_j^I$ using

$$\tilde{\mathbf{w}}_j^I = \underset{\mathbf{w}_j^I, \alpha_j^I, \xi_j^i}{\arg\min} \quad \left[\frac{1}{2}\left\|\mathbf{w}_j^I\right\| + C\sum_{i=1}^{I} u_j^i \xi_j^i\right] s.t. \quad y_j^i\left(\mathbf{w}_j^I * \mathbf{x}_j^i - \alpha_j^I\right) \geq 1 - \xi_j^i, \quad \xi_j^i \geq 0, \quad i = 1,2,\ldots,I$$

Step 3.3:

Use simple margin to select a subset of trial profiles from all unlabeled profiles, keep the S profiles with the smallest margins measured by $m_j^g = \left|\mathbf{w}_j^I \mathbf{x}_j^g - \alpha_j^I\right|$

Step 3.4:

Select the next question by selecting $\min_{s=1,\dots,S}\left[\max\left(\frac{m'^{s+}_j}{m'^{s-}_j},\frac{m'^{s-}_j}{m'^{s+}_j}\right)\right]$ among S trial profiles, where m'^{s+}_j and m'^{s-}_j are obtained by adding the s^{th} trial profile to the training data by assuming label 1 and -1, respectively.

End loop *q1*, and calculate $\tilde{\mathbf{w}}^I_j$.

Step 4:

Use $\tilde{\mathbf{w}}^I_j$ to form the candidate pool of M_j profiles for the choice questions.

Step 5:

For $q2 = 1{:}Q2$,

Step 5.1:

u^i_j and u^k_j are obtained using

$$u^i_j = \begin{cases} 1-\dfrac{\left\|\mathbf{x}^i_j-\bar{\mathbf{x}}^{I'}_{j+}\right\|}{r^{I'}_{j+}+\eta} & if\ y^i_j=1 \\ 1-\dfrac{\left\|\mathbf{x}^i_j-\bar{\mathbf{x}}^{I'}_{j-}\right\|}{r^{I'}_{j-}+\eta} & if\ y^i_j=-1 \end{cases}$$ If compared with no choice.

$$u^k_j = \begin{cases} 1-\dfrac{\left\|\mathbf{x}^{kAB}_j-\bar{\mathbf{x}}^{K'}_{j+}\right\|}{r^{K'}_{j+}+\eta} & if\ y^{kAB}_j=1 \\ 1-\dfrac{\left\|\mathbf{x}^{kAB}_j-\bar{\mathbf{x}}^{K'}_{j-}\right\|}{r^{K'}_{j-}+\eta} & if\ y^{kAB}_j=-1 \end{cases}$$ If compared with the other profile.

The class enters $\bar{X}^{I'}_{j+}, \bar{X}^{I'}_{j-}, \bar{X}^{K'}_{j+}, \bar{X}^{K'}_{j-}$ and radius $r^{I'}_{j+}, r^{I'}_{j-}, r^{K'}_{j+}, r^{K'}_{j-}$ are computed similarly as in step 3.1.

Step 5.2:

Calculate $\tilde{\mathbf{w}}^T_j$ using

$$\tilde{\mathbf{w}}^T_j = \underset{\mathbf{w}^T_j,\alpha^T_j,\xi^i_j,\xi^{kAB}_j}{\arg\min}\left[\frac{1}{2}\left\|\mathbf{w}^T_j\right\|+C\left(\sum_{i=1}^{I}u^i_j\xi^i_j+\sum_{k=1}^{K'}u^k_j\xi^{kAB}_j\right)\right]$$

$$\begin{aligned} s.t.\quad & y^i_j\left(\mathbf{w}^T_j*\mathbf{x}^i_j-\alpha^T_j\right)\geq 1-\xi^i_j \\ & y^{kAB}_j\left(\mathbf{w}^T_j*\mathbf{x}^{kAB}_j\right)\geq 1-\xi^{kAB}_j \\ & \xi^i_j,\xi^{kAB}_j \geq 0 \end{aligned}$$

Step 5.3:

Use simple margin to select the next set of S trial profile pairs (i.e., each profile pair contains profile g_1 and profile g_2) for all unlabeled choice sets, using

$\min\{\max(m^{g1,g2}_j, m^{g1}_j, m^{g2}_j)\}$, where $m^{g1,g2}_j = \left|\mathbf{w}^T_j*(\mathbf{x}_{g1}-\mathbf{x}_{g2})\right|$, $m^{g1}_j = \left|\mathbf{w}^T_j*\mathbf{x}_{g1}-\alpha^T_j\right|$, and $m^{g2}_j = \left|\mathbf{w}^T_j*\mathbf{x}_{g2}-\alpha^T_j\right|$

Step 5.4:

Select the next choice question by selecting

$$\min_{s=1,\ldots,S}\left[\max\left(\max\left(\frac{m'^{g1,g2+}_j}{m'^{g1,g2-}_j},\frac{m'^{g1,g2-}_j}{m'^{g1,g2+}_j}\right),\max\left(\frac{m'^{g1+}_j}{m'^{g1-}_j},\frac{m'^{g1-}_j}{m'^{g1+}_j}\right),\max\left(\frac{m'^{g2+}_j}{m'^{g2-}_j},\frac{m'^{g2-}_j}{m'^{g2+}_j}\right)\right)\right]$$

among the S trial profile pairs.

End loop *q2*, and calculate $\tilde{\mathbf{w}}_j^T$.

End loop *j*

Web Appendix B:Screenshots from Computer Tablet Study

Figure B1. Example of Self-Configuration Task: Computer Tablet Study

Please select the tablet you'd be most likely to purchase. For each feature, select your preferred level.

Feature	Select Feature	Cost for Feature
Brand	Apple Samsung Google Amazon Microsoft Asus Acer Lenovo Barnes&Noble	$ 0
Camera	No Camera No Rear, 0.3MP Front, 720p HD Video (+ $10) 1MP Rear, 0.3MP Front, 720p HD Video (+ $20) 3.2MP Rear, 1.2MP Front, 720p HD Video (+ $30) 5MP Rear, 1.2MP Front, 1080p HD Video (+ $40) 5MP Rear, 1.9 MP Front, 1080p HD Video (+ $50) 8MP Rear, 2MP Front, 1080p HD Video (+ $60)	$ 40
Screen Size	7 inches 8 inches 9 inches 9.7 inches 10 inches 10.6 inches	$ 0
Thickness	0.5 inch 0.41 inch (+ $10) 0.37 inch (+ $20) 0.34 inch (+ $30) 0.28 inch (+ $40)	$ 30
Weight	2 lbs 1.44 lbs (+ $20) 1.33 lbs (+ $30) 1.12 lbs (+ $40) 1 lb (+ $50) 0.76 lb (+ $60) 0.68 lb (+ $70)	$ 50
Processor Speed	1.0 GHz 1.0 GHz Dual Core (+ $10) 1.5 GHz Dual Core (+ $20) 1.5 GHz Quad Core (+ $40)	$ 40
Screen Display	132 ppi 150 ppi (+ $5) 163 ppi (+ $10) 189 ppi (+ $20) 210 ppi (+ $30) 224 ppi (+ $40) 264 ppi (+ $50) 300 ppi (+ $60)	$ 20
Storage/Memory	8 GB 16 GB (+ $20) 32 GB (+ $50) 64 GB (+ $100) 128 GB (+ $200)	$ 50

(continue to next page)

(continue from previous page)

Network/Voice Capability	WiFi WiFi, 3G (+ $50) WiFi, 4G LTE (+ $100) WiFi, 4G LTE, SIM (+ $120)	$ 100
External Connectivity	Bluetooth Bluetooth, USB (+ $20) Bluetooth, HDMI, USB (+ $30)	$ 30
Number of Apps for Download	450,000 700,000 (+ $50) 900,000 (+ $100)	$ 50
Battery Life	5 Hours 7 Hours (+ $10) 8 Hours (+ $20) 9 Hours (+ $30) 10 Hours (+ $40) 11 Hours (+ $50) 12 Hours (+ $60) 13 Hours (+ $70)	$ 40
Warranty	1-Year Warranty 2-Year Extended Warranty (+ $20)	$ 0
	Total	$ 550

Next

Figure B2. Example of "Must-Have" and "Unacceptable" Page: Computer Tablet Study

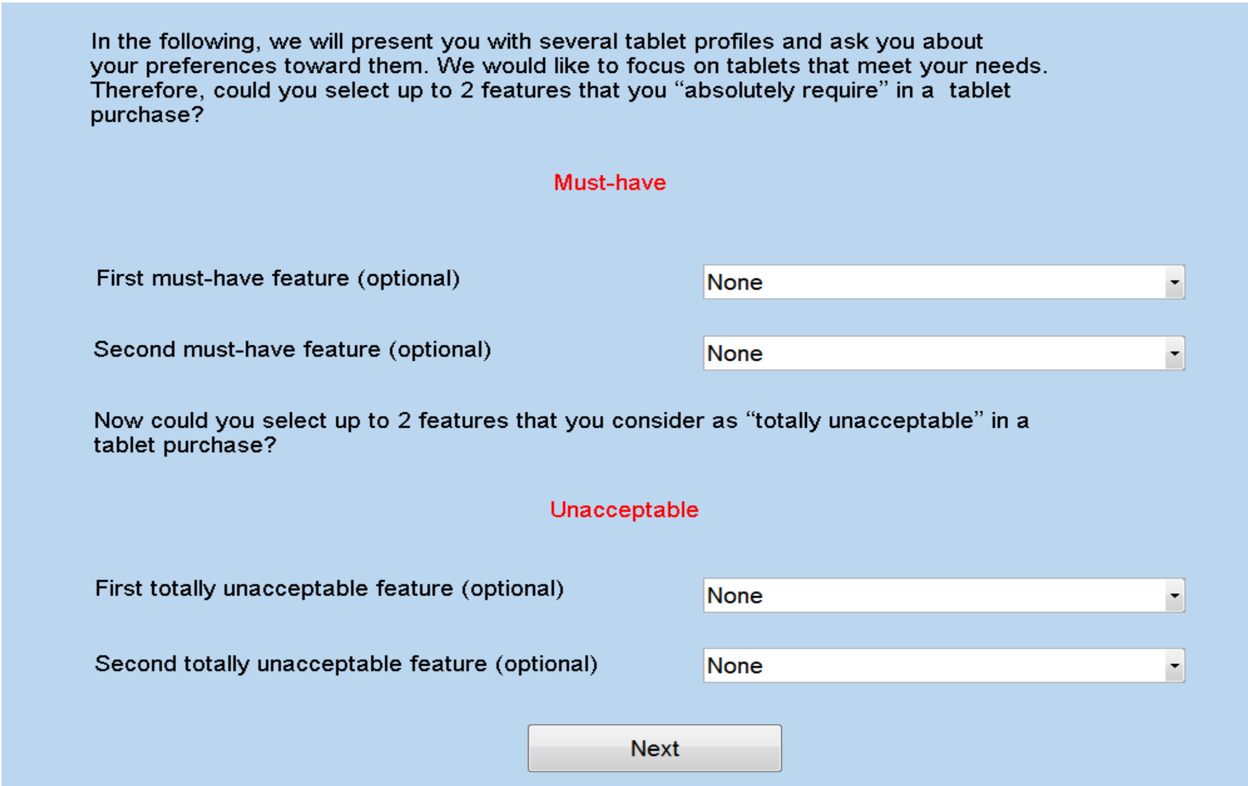
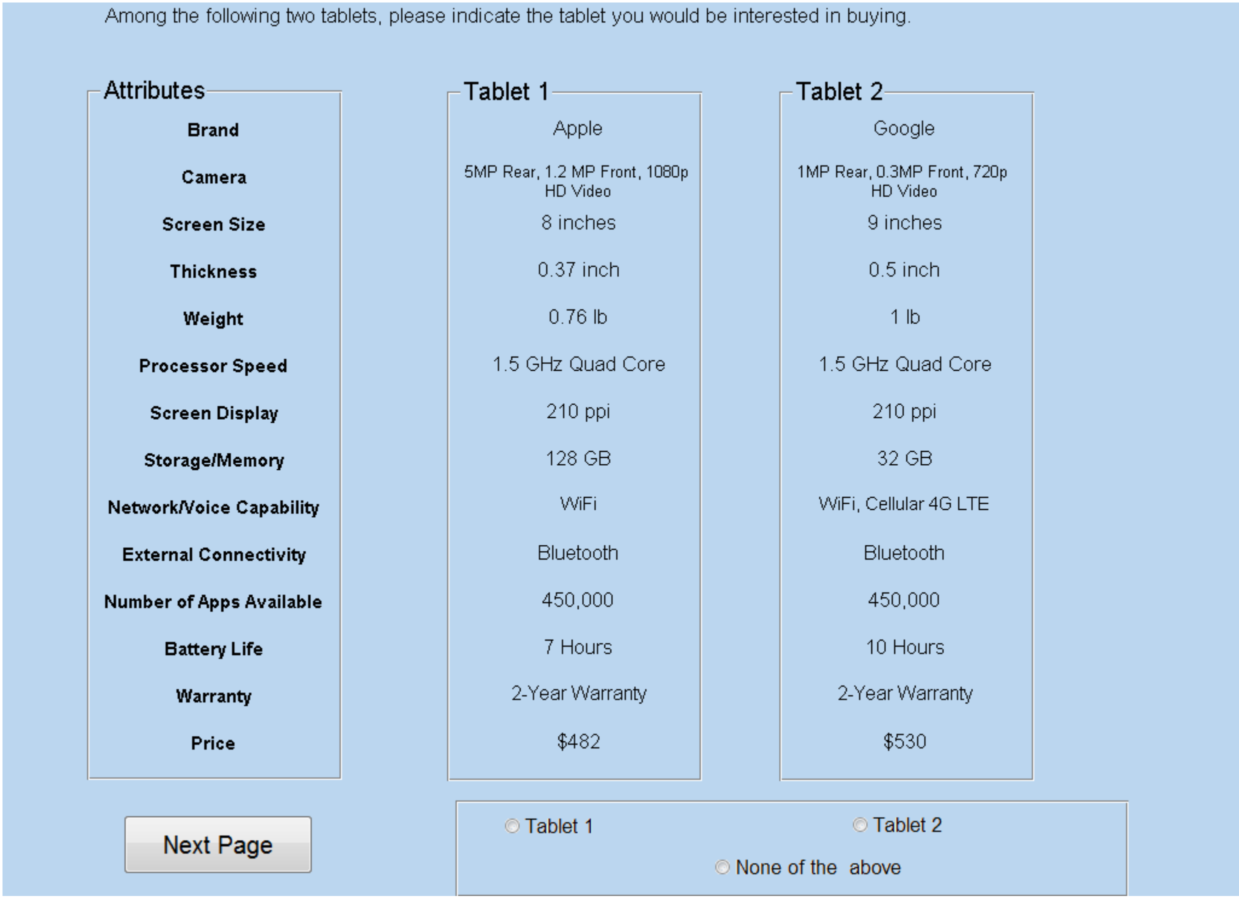


Figure B3. Example of Consideration Screening Task: Computer Tablet Study

Here are a few tablets you might like. For each one, please indicate whether it is a possibility or not.

Attributes	Tablet 1	Tablet 2	Tablet 3	Tablet 4	Tablet 5
Brand	Microsoft	Apple	Acer	Apple	Barnes&Noble
Camera	5MP Rear, 1.2 MP Front, 1080p HD Video	5MP Rear, 1.2 MP Front, 1080p HD Video	8MP Rear, 2MP Front, 1080p HD Video	5MP Rear, 1.2 MP Front, 1080p HD Video	8MP Rear, 2MP Front, 1080p HD Video
Screen Size	8 inches	8 inches	9 inches	8 inches	9 inches
Thickness	0.28 inch	0.34 inch	0.41 inch	0.34 inch	0.28 inch
Weight	1 lb	0.76 lb	1.44 lbs	1 lb	2 lbs
Processor Speed	1.5 GHz Quad Core	1.0 GHz Dual Core	1.0 GHz Dual Core	1.0 GHz	1.0 GHz Dual Core
Screen Display	264 ppi	189 ppi	224 ppi	300 ppi	300 ppi
Storage/Memory	128 GB	32 GB	128 GB	64 GB	32 GB
Network/Voice Capability	WiFi, Cellular 3G	WiFi, Cellular 4G LTE	WiFi	WiFi, Cellular 4G LTE	WiFi, Cellular 4G LTE, SIM Card Slot
External Connectivity	Bluetooth, USB, HDMI	Bluetooth, USB, HDMI	Bluetooth, USB	Bluetooth	Bluetooth, USB
Number of Apps Available	450,000	450,000	900,000	900,000	450,000
Battery Life	10 Hours	7 Hours	11 Hours	8 Hours	13 Hours
Warranty	2-Year Warranty	1-Year Warranty	2-Year Warranty	2-Year Warranty	2-Year Warranty
Price	$538	$462	$722	$537	$685
Next Page	A possibility / Won't work for me	A possibility / Won't work for me	A possibility / Won't work for me	A possibility / Won't work for me	A possibility / Won't work for me

Figure B4. Example of Decompositional Choice Stage: Computer Tablet Study



Web Appendix C: Alternative Approaches to Construct Fuzzy Membership Probability

We have explored alternative approaches to define a profile's class membership probability. One such alternative is to define a labeled profile's class membership as a function of its distances to both its own class center and the center of the opposite class. We use a method similar to Wu and Yap (2006) to develop such membership probabilities. Under this method, a high membership probability will be assigned if a profile is close to own class center and far away from the center of the opposite class. With this alternative weighting scheme, if the respondent incurs a response error, the negative effect of such an error will be penalized to a greater extent than the method we use. Nevertheless, compared to our method, this alternative weighting scheme will also incur a greater efficiency loss for correctly classified support vectors, because responses to such profiles will be given even smaller weights due to their relative proximity to the opposite class center. Our synthetic data experiment indicates that, because of the above tradeoff, this alternative weighting scheme does not provide improvement in model performance.

We have also considered the possibility of imposing an error probability distribution assumption similar to the Logit, Probit, or Gaussian Mixture models to the training data and estimating the fuzzy memberships based on such error probability distribution. Under these alternative settings, the underlying assumption is that the respondent's utility function consists of both a deterministic and a random component, with the latter following a double exponential, normal, or Gaussian mixture type of error distributions. A response error takes place if the random component in the utility function triggers a label switch in the respondent's response.

It is important to point out that, within our context of adaptive question design, standard models such as Logit and Probit cannot be used for model estimation on the fly. Within our context, the number of data points in the training data (i.e., the number of labeled profiles) is often less than the number of parameters to be estimated (i.e., the dimension of the part-worths vector), particularly at the beginning of the adaptive question study. Therefore, it is not feasible to use either Logit or Probit model to estimate the individual-specific part-worths on the fly.

To work around this problem, in our synthetic data experiments with Logit and Probit error assumptions, we use the fuzzy SVM estimation algorithm to estimate the part-worths vector on the fly. And such part-worths estimates constitute the deterministic component of the utility function. We then follow Lin and Wang (2004) to formulate the fuzzy membership probabilities under the corresponding error distributions. We find that the overall model performance in both parameter recovery and predictive validity does not change under this alternative approach to construct fuzzy membership probability. One possibility is that, while the respondent's utility function follows a Logit or Probit distribution under this setup, because adaptive question selection employs selective rather than random sampling, the resulting training data may not necessarily mirror the initial underlying distribution. As a result, the performance of our distribution-free fuzzy membership estimation is nearly equivalent to that from a fuzzy membership probability estimated directly based on Logit or Probit error distributions.

We have also considered the possibility of estimating the fuzzy membership probability based on Gaussian Mixture models. The underlying assumption is that response errors in the labeled data may come from two separate Gaussian distributions. And the respondent's response error is affected by a mixture of the two distributions. Gaussian Mixture models are well posed when there are sufficient data points in the training data. Similar to the issue faced by Logit and Probit model discussed earlier, it is not feasible to use Gaussian Mixture models to estimate the individual-specific part-worths on the fly, as the number of labeled profiles can be less than the dimension of the part-worths vector in our context. Moreover, we also do not have sufficient information to estimate the mixture weights with very few data points. Our conjecture is that such models may not be directly applicable in our particular context of adaptive question design.

Web Appendix D: Supplementary Information on Empirical Application

Table A1. Attributes and Attribute Levels Used in the Digital Camera Study

Attributes	*Levels*
Brand	Canon, Panasonic, Nikon, Olympus, Sony
Image Stabilizer	None, Electronic Stabilizer, Optical Stabilizer
Camera Size	Pocket Size, Medium Size, Large Size
LCD Size	2.5 inches, 3 inches, 3.5 inches
Life of Battery	150 pictures, 300 pictures, 450 pictures, 600 pictures
Optical Zoom	3x, 6x, 9x, 12x
Resolution	8 Mega Pixels, 10 Mega Pixels, 12 Mega Pixels, 14 Mega Pixels
Video Clip	Standard Video, HD Video
Warranty	1 year, 2 years, 3 years
Price	Summed price with a base of $100 and a +/- 30% random variation

Table A2. Comparison of Feedback Questions: Digital Camera Study

Method (Sample Size)	*Feedback Q1*		*Feedback Q2*		*Feedback Q3*	
	Ave.	*SE*	*Ave.*	*SE*	*Ave.*	*SE*
Cond 1: Proposed Method (N=73)	4.014*	.100	3.671*	.151	3.726*	.112
Cond 2: FSVM without Collaborative Filtering (N=69)	3.838*	.102	3.250	.128	3.515*	.144
Cond 3: Self-Explicated (N=66)	3.621	.118	3.303	.154	3.667*	.132
Cond 4: Upgrading (N=80)	2.313	.106	2.625	.124	2.363	.113
Cond 5: ACBC (N=75)	1.800	.093	2.267	.119	2.373	.128
Cond 6: CBC (N=62)	2.048	.114	2.387	.127	2.629	.144

*best in column or not significantly different from best in column at the .05 level

Q1-Q3 are 5-point scale questions with 1 being "strongly disagree" and 5 being "strongly agree":

- Q1: The camera configuration I was asked to evaluate seem realistic.
- Q2: The survey format made it easy for me to give realistic answers that reflect exactly what I will do if buying a real digital camera.
- Q3: The way the cameras were presented made me want to slow down and make careful choices.

Table A3. Comparison of Survey Duration: Digital Camera Study

Method (Sample Size)	*Survey Duration (in minutes)*	
	Ave.	*SE*
Cond 1: Proposed Method (N=73)	11.6	.481
Cond 2: FSVM without Collaborative Filtering (N=69)	10.8	.494
Cond 3: Self-Explicated (N=66)	6.0	.223
Cond 4: Upgrading (N=80)	6.5	.281
Cond 5: ACBC (N=75)	10.9	.111
Cond 6: CBC (N=62)	13.8	.193

Table A4: Comparison of Survey Duration: Computer Tablet Study

Method (Sample Size)	*Warm-up and Initial Validation (in minutes)*		*Preference Elicitation (in minutes)*	
	Ave.	*SE*	*Ave.*	*SE*
Cond 1: Proposed Method (N=35)	6.1	.470	12.4	.646
Cond 2: FSVM without Consideration Screening (N=36)	5.8	.479	8.6	.466
Cond 3: Self-Explicated (N=39)	7.1	.610	8.1	.401
Cond 4: ACBC (N=41)	5.6	.401	13.2	.554

Web Appendix E: Supplementary Information on Synthetic Data Experiments

As discussed earlier, in the context of complex products, it is unfeasible to evaluate all profiles using adaptive active learning without excessive delays in between questions. When the focal method is used, a candidate pool of 20,000 product profiles is used for each synthetic respondent in the adaptive question selection. When the method by Dzyabura and Hauser (2011) is used, we follow Dzyabura and Hauser (2011) by employing uncertainty sampling in which the 400 most uncertain profiles are evaluated for each subsequent question. We choose 400 rather than 1,000 most uncertain profiles as in Dzyabura and Hauser (2011) to speed up computational speed. As shown in Table 4, both sampling methods are sufficient in recovering the true conjunctive decision rules. All results also hold qualitatively when we increase the size of the uncertain sampling from 400 to 1,000.

References

Dzyabura D, Hauser JR (2011) Active Machine Learning for Consideration Heuristics. *Marketing Sci.* 30(5):801-819.

Lin CF,Wang SD (2004) Training Algorithms for Fuzzy Support Vector Machines with Noisy Data. *Pattern Recognition Let.* 25: 1647-1656.

Wu K, Yap KH (2006) Fuzzy SVM for Content-Based Image Retrieval. *IEEE Computational Intelligence Mag*. 1(2): 10-16.