

Online Footsteps to Purchase: Exploring Consumer Behaviors on Online Shopping Sites

Munyoung Lee*, Taehoon Ha*, Jinyoung Han**,
Jong-Youn Rha*, and Ted "Taekyoung" Kwon*
(mylee@mmlab.snu.ac.kr)

Seoul National University*
University of California, Davis**
2015. 6. 30.



Outline

- Motivation
- Datasets
- Methodology
- Analysis of online shopping sites
- Analysis of behaviors towards purchase
- Prediction of purchase
- Conclusion

Motivation (1/2)

- Online markets
 - Important part of the Internet economy
 - Gaining much interest in research community and industry
 - e.g., study motivations of online shoppers, consumer behaviors...
- *What drives consumers to purchase products* on online markets?
 - It has not been thoroughly investigated due to lack of log data
 - e.g., consumers' online behaviors including their purchases
- Our work moves forward from prior studies
 - By analyzing consumers' actual online behaviors that lead to actual purchases
 - By using datasets from multiple online shopping sites

Motivation (2/2)

- Research questions?
 - What are the similarities and differences in the consumer behaviors across four different online sites?
 - What are the distinct behavioral patterns of consumers who actually purchase items?
 - Can we predict whether a consumer will make a purchase for an item or not?
 - Previous online activities
 - Items properties
- To answer the questions, we analyze the massive log data obtained from four popular online shopping sites

Datasets

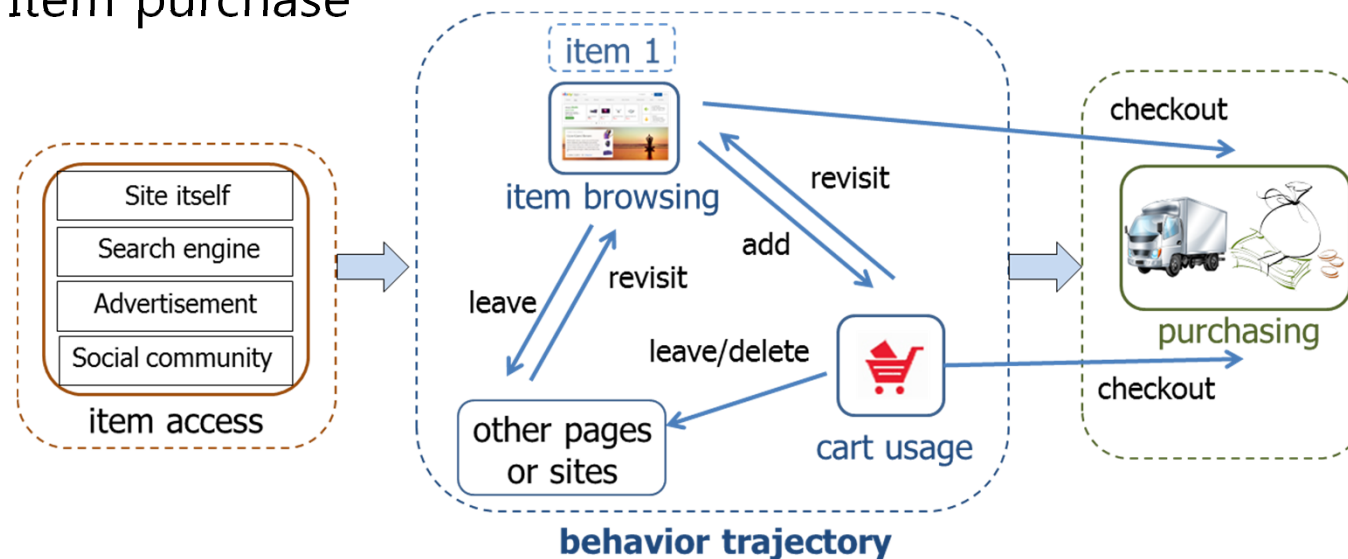
- Consumers' behavior logs in online markets
 - Four online shopping sites in South Korea
 - General-purpose marketplace: S1 (like eBay)
 - Special-purpose shopping sites: S2 (clothes), S3 (home deco), S4 (sports)
 - Web access histories of 0.7M consumers
 - All users visiting the four sites (using 3rd party cookies)
 - Period: 24 days (Mar. 1, 2014 ~ Mar. 24, 2014)
 - More than 100 million consumer activities
 - Request time, cookie, IP address, URL, item title, category, price...
- Information of four online sites

site	Category	# logs	# consumers	Total sales	# items
S1	General	76,375,439	564,786	\$59.9M	1,808,893
S2	Clothes	16,836,901	64,467	\$15.5M	50,492
S3	Home decoration	5,062,218	45,271	\$6.9M	23,988
S4	Sports	3,348,474	17,229	\$2.3M	1,681

* Upon requests from the dataset provider, we make anonymized the names of the four online sites

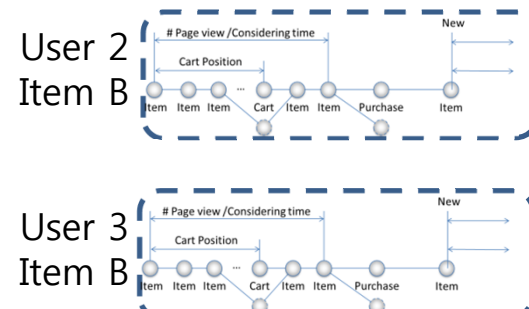
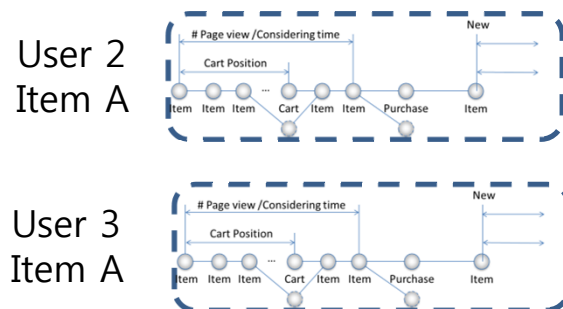
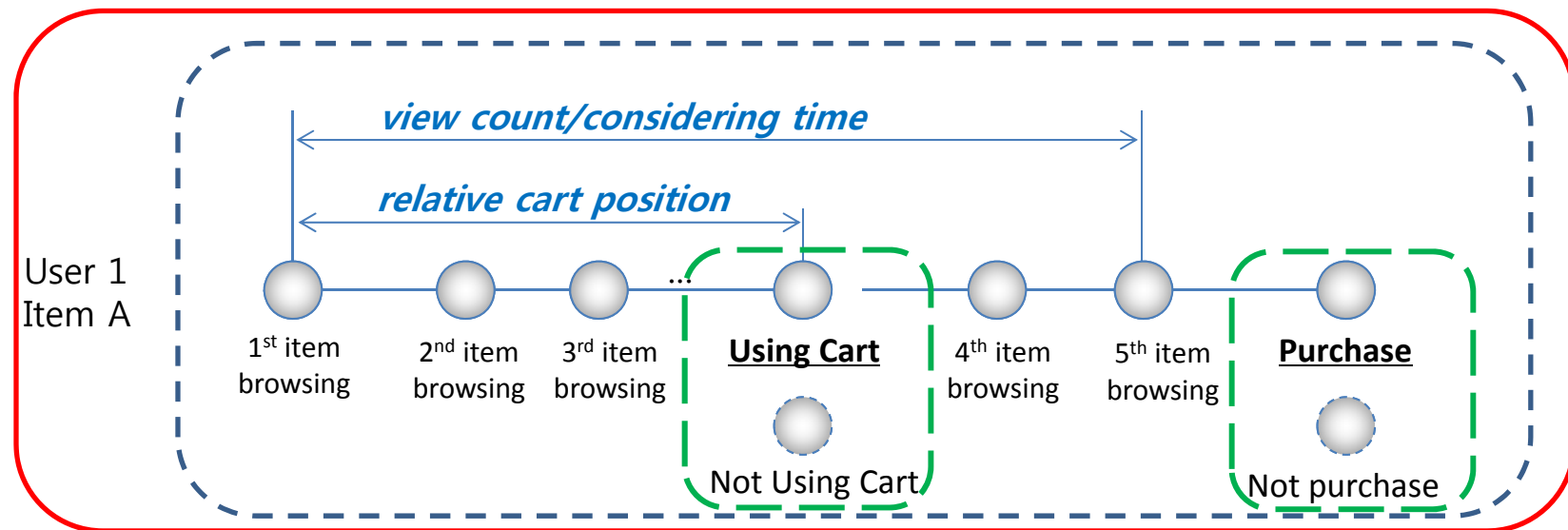
Methodology (1/2)

- Analyzing consumer's online shopping process
 - Item access
 - How a user comes to visit the item page
 - Behavior trajectory
 - A sequence of consumer's behaviors for each item
 - Item purchase



Methodology (2/2)

- Data processing
 - Constructing **per-item behavior trajectory** of each consumer on the sites



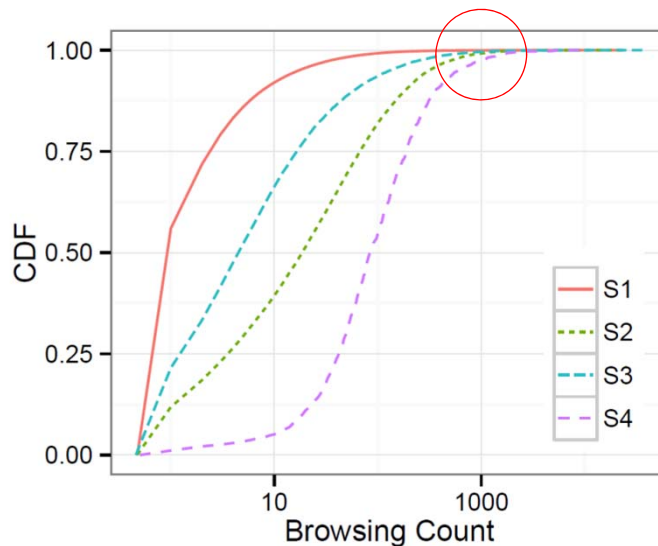
User	Item	Properties
1	A	...
2	B	...

Analysis overview

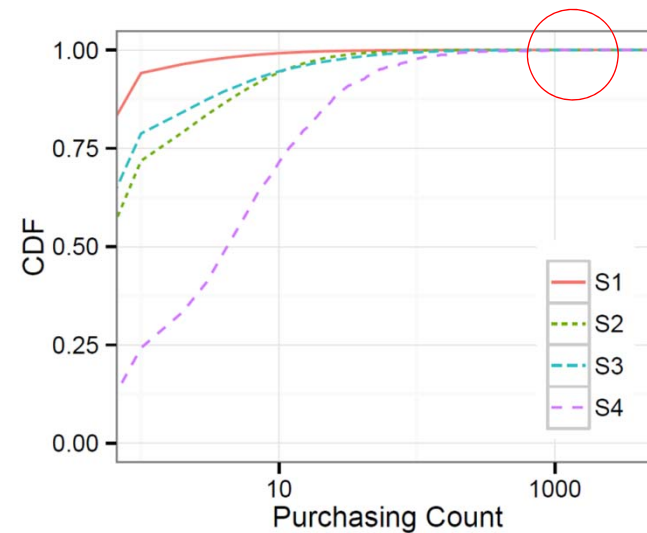
- Analysis of online shopping sites
 - Examining item properties by sites and on-site consumer behaviors
 - Item popularity (ranking)
 - Source of item accesses (referral site for a given item page)
- Analysis of behaviors towards purchases
 - Exploring consumer's actions that lead to actual purchases
 - Item browsing (viewcount, consideration time)
 - Cart usage (using cart, relative cart position)
- Prediction of consumers' purchases
 - Using the features associated with the behaviors towards purchases

Analysis of online shopping sites (1/2)

- Item popularity
 - Focus on on-site consumer behaviors
 - Browsing behavior (browsing items),
 - Purchasing behavior (purchasing items)
 - Skewness of item popularity
 - Online behaviors are different depending on the shopping sites
 - Only a small number of items are browsed and purchased more than 1,000 times by online consumers



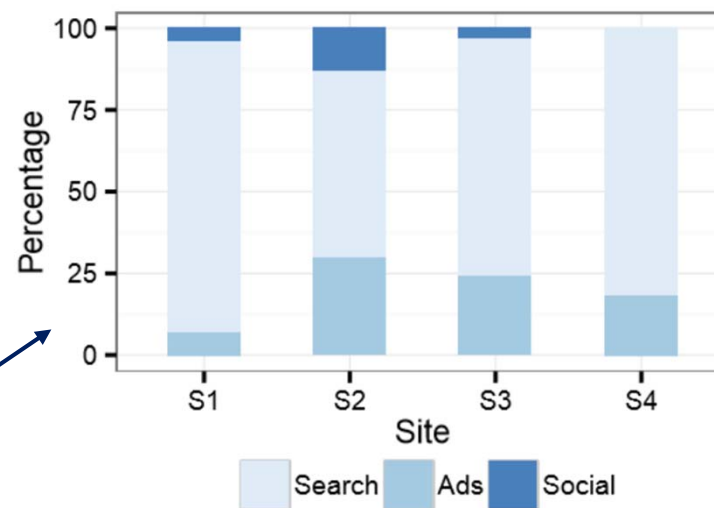
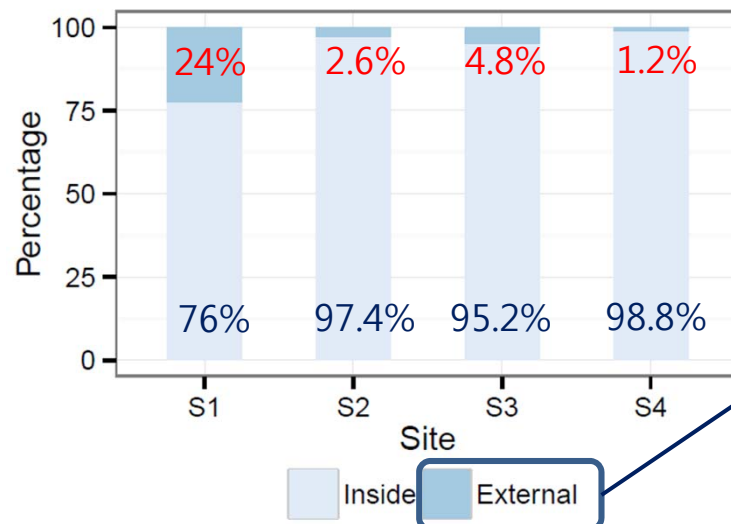
* Browsing count: how many times an item is browsed



* Purchasing count: how many times an item is purchased

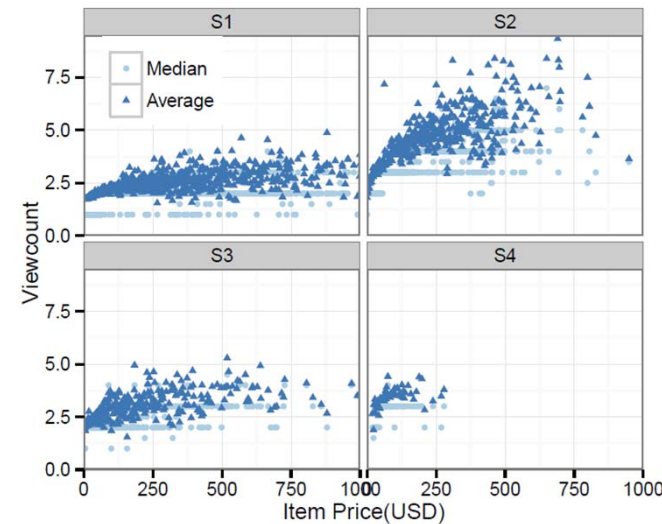
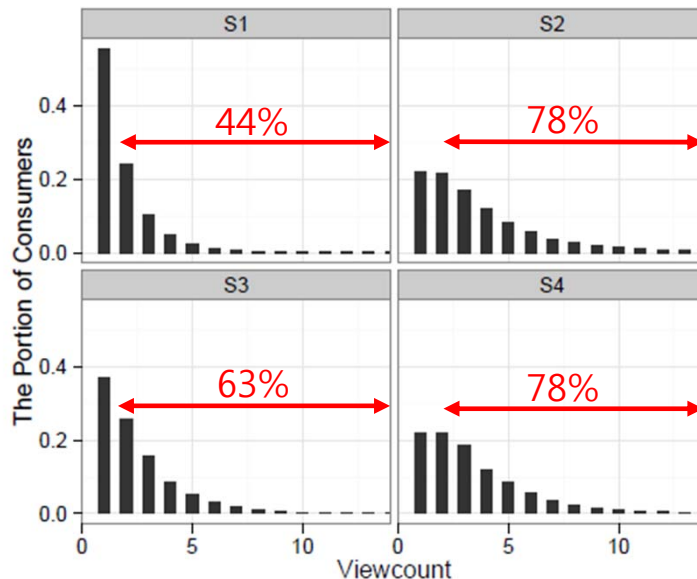
Analysis of online shopping sites (2/2)

- Source of item accesses
 - Investigating *how consumers come to access items*
 - By examining consumer's previous URL (referral site) when the consumer visits the item page for the first time
 - Most of consumers visit item pages **from the online shopping sites themselves**
 - e.g., searching within the sites or browsing sub menus
 - Significant differences between the general-purpose marketplace S1 (like eBay) and the other special-purpose shopping sites



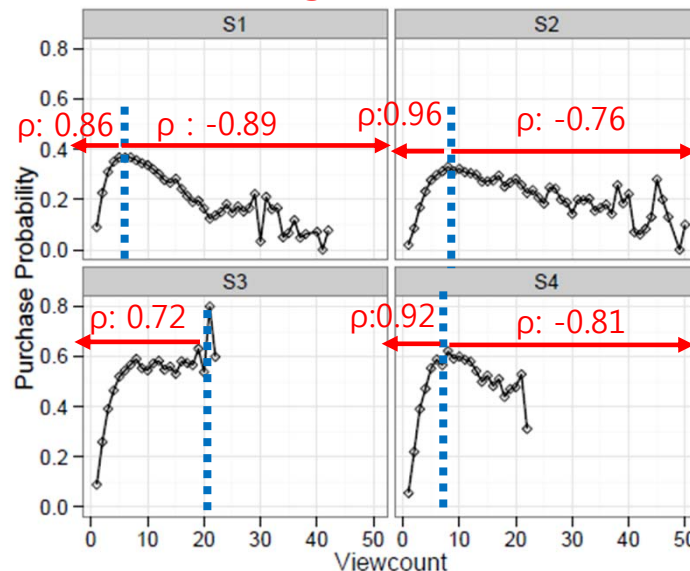
Behaviors towards purchases (1/4)

- Item browsing (cases of purchases)
 - Q: how the browsing behaviors of consumers lead to their purchases?
 - Viewcount: how many times a user has viewed the given item
 - Analyzing the viewcount of purchased items
 - Consumers who purchased items are shown to **browse the same item two times or more** before purchasing (44%~78%)
 - Consumers tend to browse the same item more as its price increases (weak positive correlation)



Behaviors towards purchase (2/4)

- Item browsing
 - Calculating the purchase probability against the viewcount
 - Purchase probability = $\frac{\text{\# of purchased items on a particular condition}}{\text{\# of total items on a particular condition}}$
 - The growth of the viewcount up to a particular point (called *view threshold*) results in the increase of the purchase probability
 - There is a **strong positive correlation** between purchase probability and viewcount until the view threshold
 - Implication
 - **Browsing behaviors** can be used as a factor to predict purchases



Pearson correlation

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

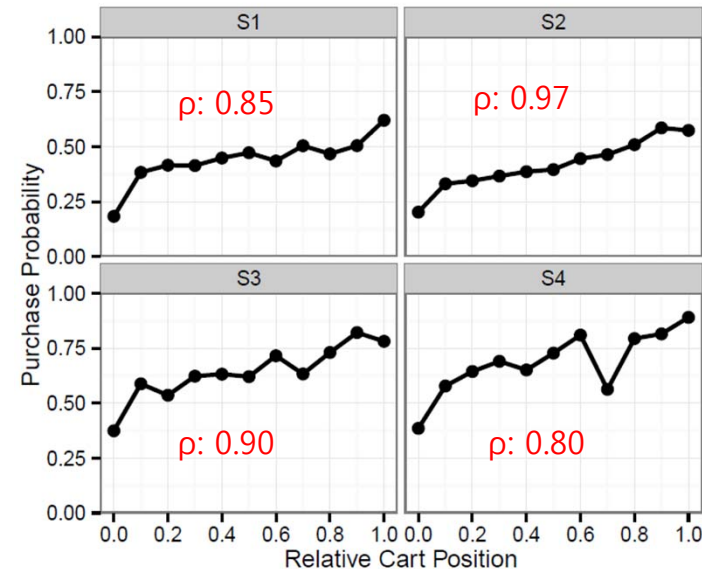
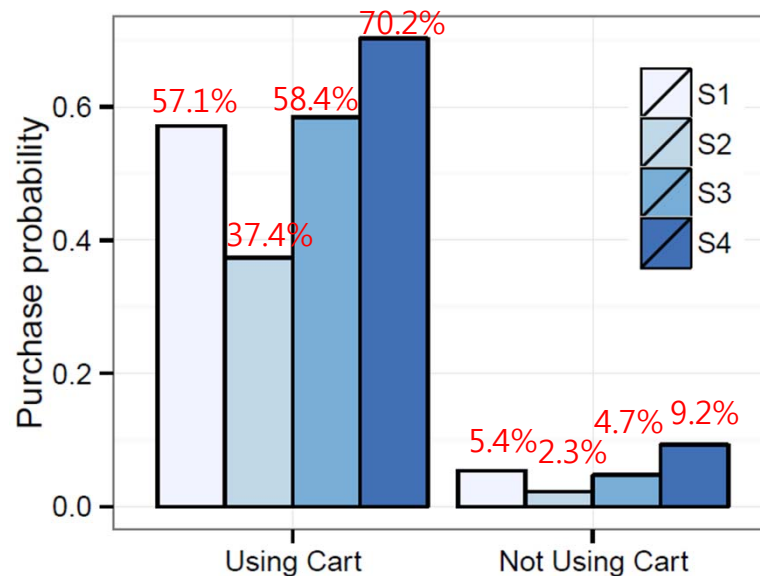
Behaviors towards purchase (3/4)

- Cart usage
 - Q: how many consumers utilize online carts for their shopping?
 - Examining consumers whose carts are used at least once
 - A majority of consumers in the four sites utilize carts (43~73%)
 - e.g., more than 70% of users in sites S2 and S3 utilize carts when they do the shopping for fashion items and home deco items

site	# of consumers	# of cart users	ratio
S1 (marketplace)	564,786	324,356	57%
S2 (fashion)	64,467	45,253	70%
S3 (home deco)	45,271	33,236	73%
S4 (sports)	17,229	7,409	43%

Behaviors towards purchase (4/4)

- Cart usage
 - Q: how the user's cart usage behavior leads to purchases?
 - Comparing purchase probability against cart usage (using Cart)
 - Ratio of purchased items among items placed in the cart
 - Ratio of purchased items among items not placed in the cart
 - Calculating relative cart position
 - relative cart position = $\frac{\text{viewcount before placing into a cart}}{\text{total viewcount}}$
 - Implication
 - cart usage behaviors can be used as a feature to predict purchases



Later put in
cart, higher
purchase
probability

Prediction of purchase (1/2)

- Problem definition
 - To identify whether a user will purchase an item or not based on the observed features
- Feature set
 - Item: price, item popularity, total # of purchases
 - Behavior: viewcount, consideration time, cart usage, relative cart position
 - All: Item features + Behavior features
- Classifier
 - Support Vector Machine (SVM) classifier
 - Performance metrics
 - Precision, recall, F1 score, accuracy, and area under the ROC curve (AUC)
- Validation
 - 5-fold cross-validation

Prediction of purchase (2/2)

- Model using Behavior features
 - Higher accuracy (i.e., from 0.80 to 0.84) than the Item model across four online shopping sites
 - Comparable performance with the model using All features
- Behavior features are **good predictors** for actual purchases

		S1 (General)					S2 (Clothes)				
		Precision	Recall	F1 score	Accuracy	AUC	Precision	Recall	F1 score	Accuracy	AUC
Item	Not buy	0.56	0.87	0.68	0.60	0.67	0.64	0.37	0.47	0.58	0.58
	Buy	0.72	0.32	0.45			0.55	0.79	0.65		
Behavior	Not buy	0.73	0.92	0.82	0.80	0.78	0.81	0.89	0.85	0.84	0.84
	Buy	0.90	0.67	0.77			0.88	0.79	0.83		
All	Not buy	0.74	0.92	0.82	0.80	0.79	0.81	0.89	0.85	0.84	0.84
	Buy	0.90	0.67	0.77			0.88	0.79	0.83		
		S3 (Home decoration)					S4 (Sports)				
		Precision	Recall	F1 score	Accuracy	AUC	Precision	Recall	F1 score	Accuracy	AUC
Item	Not buy	0.54	0.96	0.69	0.57	0.56	0.55	0.91	0.68	0.58	0.59
	Buy	0.79	0.17	0.29			0.74	0.27	0.39		
Behavior	Not buy	0.80	0.88	0.84	0.83	0.83	0.78	0.89	0.83	0.82	0.81
	Buy	0.86	0.78	0.82			0.87	0.75	0.81		
All	Not buy	0.80	0.88	0.84	0.83	0.83	0.79	0.90	0.84	0.83	0.82
	Buy	0.86	0.78	0.82			0.88	0.77	0.82		

Conclusion

- Comprehensive measurement study on consumer behaviors on four online shopping sites
- Findings
 - Substantial portion (24%) of consumers in a general-purpose marketplace discover items from external sources, while most (>95%) of consumers in special-purpose shopping sites directly access items from the sites themselves
 - Item browsing patterns and cart usage patterns are the important predictors of the actual purchases
 - Our proposed model achieved over 80% accuracy in predicting consumers' purchases

Q&A



Munyoung Lee
Seoul National University
mylee@mmlab.snu.ac.kr