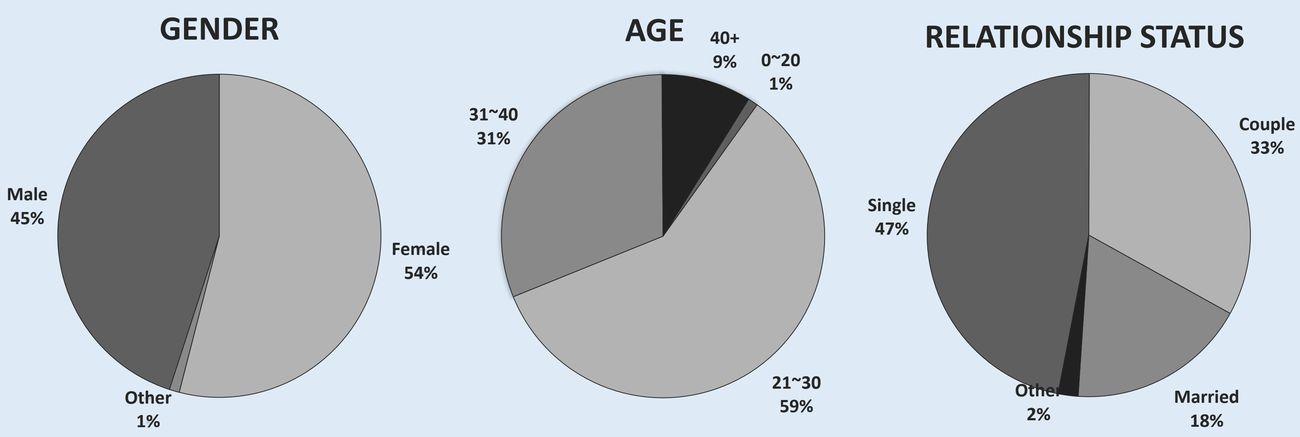
Predicting User's Online Shopping Tendency During Shopping Holidays

Cheng-You Lien, Guo-Jhen Bai, Ting-Rui Chen, Hung-Hsuan Chen
Computer Science and Information Engineering, National Central University

Abstract

The number of sales during the shopping holidays continues growing in recent years. Thus, many E-Commerce (EC) websites spend much money and effort for marketing before these shopping holidays. However, in this study we found that only part of the Internet users indeed visited the EC-websites more often than usual during the shopping holidays. Thus, the increase of the sales probably comes from few individuals. Additionally, we found that users' tendency to visit the EC websites during the shopping holiday is predictable based on simple supervised classifiers. Thus, an EC website runner can identify the potential visitors and non-visitors beforehand and apply different marketing strategy to different users.

Description of the data



We collected 517 users' complete browsing history stored in their Google Chrome browsers.

- Most of browsing histories were recorded form **August 2016** to **December 2016**.
- Users are familiar with Internet and have experience in **online shopping**.
- We define the **shopping ratio** as follow:

$$\text{shopping ratio}: \frac{number\ of\ visits\ on\ the\ shopping\ websites}{total\ number\ of\ visits}$$

The average shopping ratio on each day is shown in Figure 1.

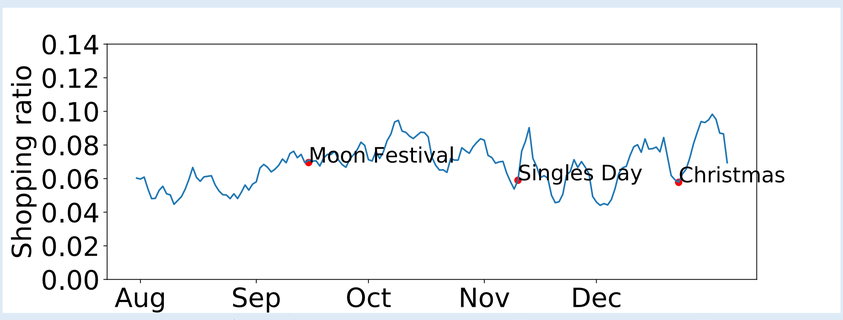


Fig. 1: Users' visiting ratio on shopping websites

Data preprocessing

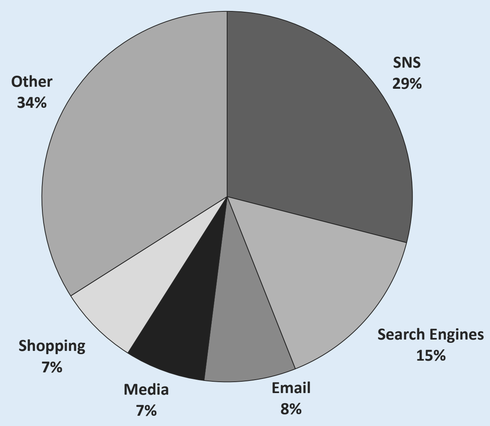


Fig. 2: The proportions of the visited website categories

- We converted all the URLs to 73 categories based on *Web Filter Lookup*, a public service to **convert URLs into categories**.
  - Because the visited URLs are highly skewed
- For each user, we compute her/his accumulated visits of each category on each day.

Experiment

- We selected three shopping holidays (**the Moon Festival**, **the Single Day** and **Christmas**) during Au-gust 2016 and December 2016.
- We selected features from users' **demographical information** (genders, ages and relationship status) and **users' browsing history**.
- We define a user as a **positive instance** if the user's average shopping ratio increases during the target shopping holiday.

	Moon Festival		Singles Day		Christmas	
	Training	Test	Training	Test	Training	Test
KNN	0.71	0.55	0.75	0.63	0.78	0.72
LR	0.73	**0.65**	0.70	0.61	0.82	0.73
SVM	0.74	0.64	0.84	**0.64**	0.84	**0.77**
RF	0.99	0.60	0.99	0.60	0.99	0.68

Table 1. The average AUCs of the training and the test datasets of different classifiers on the three holidays

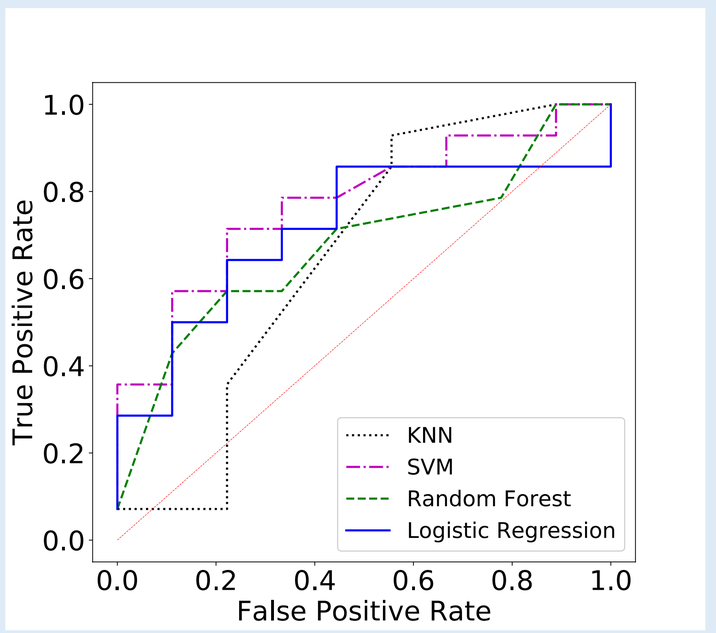


Fig 3. The ROC curve of Christmas based on the test data

Discussion and conclusion

1. The increases of sells during the shopping holidays may come from few individuals
2. Potential buyers and non-buyers can be identified based on simple supervised classifiers
   - The EC runners may create customized marketing strategies toward different types of users.
3. The AUCs increase as more training data are available