SHOP: Store Habits Of People

Sougata Sen, Dipanjan Chakraborty, Dipyaman Banerjee, Archan Misra, Nilanjan Banerjee, Vigneshwaran Subbaraju, and Sumit Mittal

Introduction

Identifying a shopper's in-store preferences, attitudes and intentions is an important objective for retailers. Some initial work has aimed to identify shoppers' activities while visiting malls [2]. Current techniques for understanding a shopper's behaviour and intent, such as shadowing and video analysis [1], do not scale well or suffer from privacy concerns. In this work we propose a system called *SHOP*, that utilizes smartphone-embedded sensors to capture and derive features from a user's in-store movement and activity patterns, and then uses the feature vector to classify the intent of the shopper.

Challenges

There are two key challenges to tackle:

- *Behaviour Diversity*: Due to a variety of demographic and individual-level traits [3], shoppers with the same intent can exhibit very different observable behavioural patterns.
- *Lack of Labelled Data*: Given our desire to build a crowd-scale system, it is simply infeasible to assume the availability of significant labelled, training data from individual shoppers.

Model Formation Methodologies

SHOP employs a novel semi-supervised learning technique which is able to encode both shopper demographics and environmental factors in its data model. It can also learn from both labelled and unlabelled data. Our main aim in this work is to:

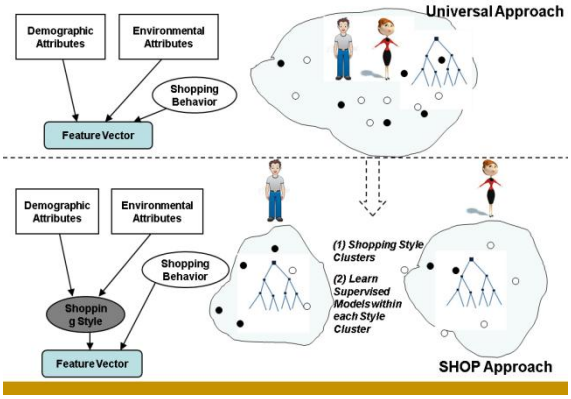
- Predict the shopping behaviour of the episodes which have not been labelled
- Evolve the model over time as new data is received.

Unlike the conventional approach which uses the demographic & environmental attributes directly (and thus cannot operate when these values are not available), we propose to generate the feature vector, using an implicit property: the *shopping style*, that is not directly observable, but is influenced by demographic attributes and environmental attributes. So if we create separate classifiers for each style, the classification will improve.

User Study and Initial Results

We conducted a user study in the food court of a shopping mall in New Delhi with 30 users. The users were given certain task to emulate the following behaviours – (i) having buying intention and focused (BI-F), (ii) having buying intention but confused (BI-C) and (iii) having no buying intentions (NBI).

We evaluated *SHOP* against 3 alternatives: (i) Single level supervised classifier (ii) Hierarchical supervised classifier and (iii) Unsupervised clustering based classifier. We found that (i) the two tier and single level classifiers have similar accuracy (ii) unsupervised clustering performs quite well when all incoming episodes are labelled and (iii) Our approach performs the best, when the data is only partially labelled.



The table below shows the classification accuracy, separated out by demographic attributes (to help us assess their impact). We find we can achieve improvement in accuracy by up to 20%.

Demographics	NBI vs BI	BI-C vs BI-F
None	71.58%	59.88%
Male	69.00%	68.50%
Female	79.83%	64.50%
Vegetarian	81.33%	56.22%
Non-vegetarian	92.50%	78.44%

On-going and Future Work

We are currently conducting a user study in a different food court to collect more shopping episodes, and thus validate/improve our system. We will also focus on defining how the model evolves over time, and how such models can be transferred from one location to another.

References

[1] Kröckel, Johannes, and Freimut Bodendorf. "Intelligent Processing of Video Streams for Visual Customer Behavior Analysis." *ICONS 2012.*

[2] Lee, SangJeong, et al. "Understanding customer malling behavior in an urban shopping mall using smartphones." Pervasive and ubiquitous computing adjunct publication. ACM, 2013.

[3] Lane, Nicholas D., et al. "Community Similarity Networks." Personal and Ubiquitous Computing: 1-14.

This work is supported in part by the Singapore Ministry of Education Academic Research Fund Tier 2 under the research grant MOE2011-T2-1-001 and conducted as part of an Open Collaboration Research project 11360571 with IBM Research. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the granting agency or Singapore Management University.