Critique of GruMon Paper

Kashev Dalmia, Ryan Freedman, Terence Nip
{dalmia3, rtfreed2, nip2}@illinois.edu

September 21, 2015

1 Summary

The paper summarizes the process that the researchers went through to develop the GruMon system. After discussing previous attempts to create classifiers with the same goal the paper describes the method it uses. The system is used to distinguish individuals from groups by detecting activity/location information using phone sensors, computing similarities between pairs, and passing them through an SVM classifier. The motivations behind developing GruMon were primarily for group detection to allow for better advertising and resource planning in crowded areas. There were three trials used for measurement and testing: a Korean mall, a Singaporean mall, and an International Airport. The three trials had varying means of detection, participants, and incentives for participation.

2 Critique

First, we found it curious that the authors chose to use a control group of phones for the more "difficult to navigate" Mall 2. It is possible that a more random sample of phones, with more run of the mill specs, wouldn't have been able to navigate the second mall as well as the Samsung Galaxy S III's. Thus, their strong precision and recall numbers are possibly inflated.

Second, the training of the classifier is problematic. They stated that they trained the classifier on ground truth from each mall. They did not state clearly how the initial classifier for the airport data was trained. Though their accuracy is quite good, it is only quite good for venues that have been pre-scouted, by willing participants. If for every new venue GruMon is deployed to, a contingent of paid participants has to be sent out with a special app on their phones, using GruMon for useful purposes is going to become extremely costly and impractical extremely fast.

Finally, though the authors say that there are issues with scaling GruMon, it is strange that the only data sets that they have with ground truth are $O(100)$, nor are there any ground truth labels for any data that's not in a mall. Without knowing the precision for venues with different usage patterns (like airports, concert venues, larger more open venues like university campuses), which might be larger than $O(100)$, it is hard to declare this problem solved by GruMon. It is possible that their classifier only works well in malls. The authors say that Mall 2's access points were poor and not as useful because of the large atrium in the mall, but no further detail is given. There is too much variability in the malls to know for certain what the causes of the discrepancies in precision are.

3 Praise

Despite our criticisms, there several things that the paper did well. At the end of the day, for both Mall data sets, GruMon was able to achieve over 90% precision, which is quite good for the purposes of advertising. Another thing that they did well was feature selection; though they did use several features for their classifier, all those features were chosen based on an insight of how actual humans behave in groups. For instance , the barometer was used and given high weight because humans in groups tend not to be on different floors of a venue, even if they're visiting different stores within the floor. It would have been easy to just log values for every sensor the phone had and throw them in a classifier, but instead, they chose carefully, and thus saved battery life of the devices used. Finally, it was interesting and impressive that they had a good method of combating poor external infrastructure, like the few and poorly placed WiFi access points in Mall 2 (though we have our concerns, as mentioned above.)