# Noise Reduction of Mobile Sensors Data in the Prediction of Demographic Attributes

Itay Hazan, Dr. Asaf Shabtai

The new, recently developed method will be able to be integrated into any app downloaded from the Google Play Store and will provide a well-founded prediction of users' gender differences from the moment the installation completes. This new innovative algorithm uses information collected by utilizing already existing sensors that are accessible to any application (such as motion sensors that detect user movements). The prediction will provide 80% accuracy, improving over time to 92%, in a completely transparent way to the app user and the application programmer. The implication is that any application can obtain the most important demographic feature of the user.

In recent years, we have witnessed a fast-growing use of smartphones and mobile devices of various types. Ease of use, along with the amount of applications and capabilities, manage to make these devices integral parts of our life.

Along with great functionality, we are witnessing the wide spread of advertising and marketing content through these devices. While the world of marketing becomes more user specific, services and content providers make an effort to personalize ads and recommendations to make them better received by users.

Facebook, for example, will offer users customized commercial content based on their profile information, such as: age, country, marital status and other relevant information updated by the users on Facebook. Couples, for example, that have updated their relationship status to "Engaged" will be shown relevant advertisements, such as wedding rings and event halls. Other learning capable websites such as YouTube will recommend videos based on users' activities and frequently viewed videos.

In order to save users from the process of revealing and explaining their preferences, inferring gender through the mobile device automatically helps replace older methods, such as text analysis. The average user would not be inconvenienced if an app would measure the device's position every so often in order to predict the user's gender, but it would be more concerning if a tracker would analyze SMS messages or browsing history—even if this is done only for statistical analysis. As evidence, we know that these methods require the user's permission and might discourage anyone from downloading the app.
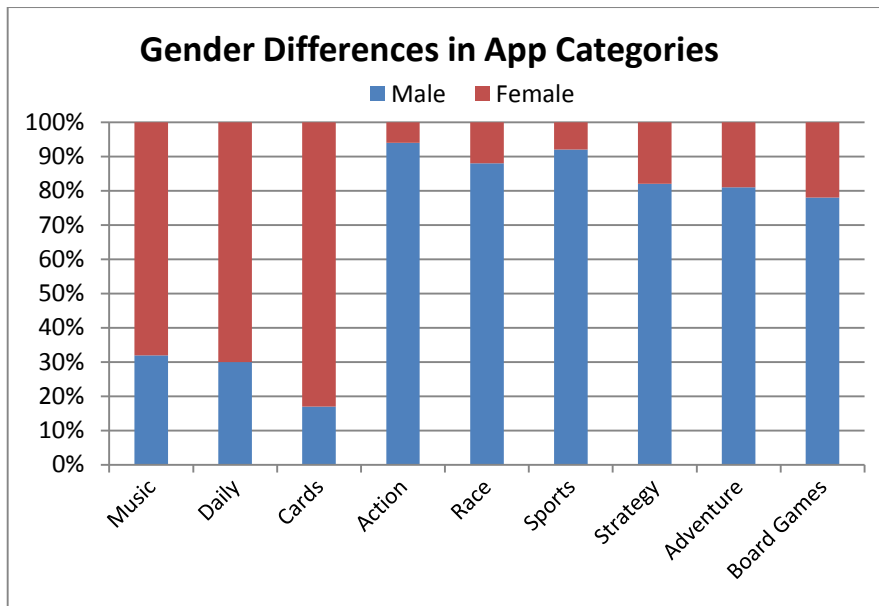
On the other hand, there are ways that do not interfere with users' privacy and still deliver the same conclusions with higher precision. For instance, by detecting the specific movement of holding the device upside down in the pants pocket and counting, it show us two different distributions of men and women as can be seen in

the graph below, where each axis counts the amount of times the user held the device in a particular way



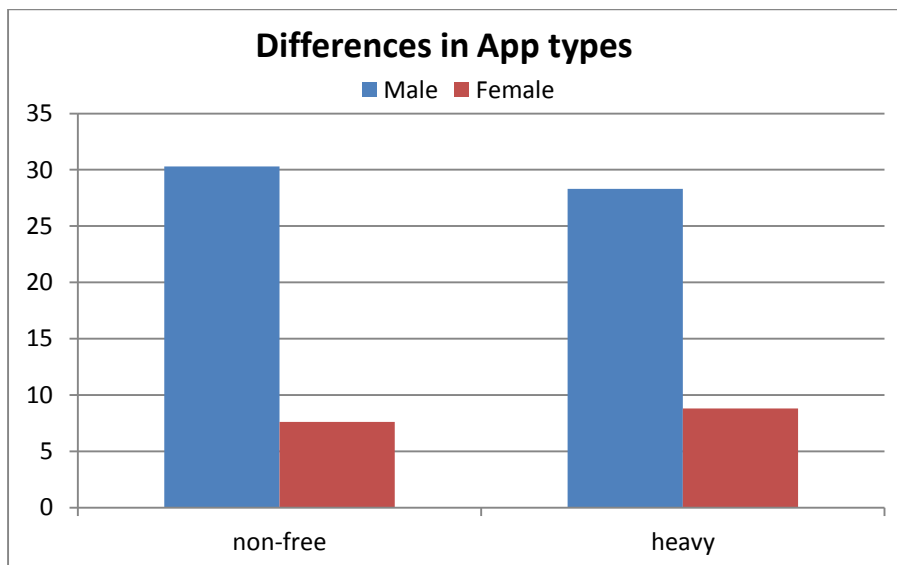Distribution of Users in the "Pocket" Postions
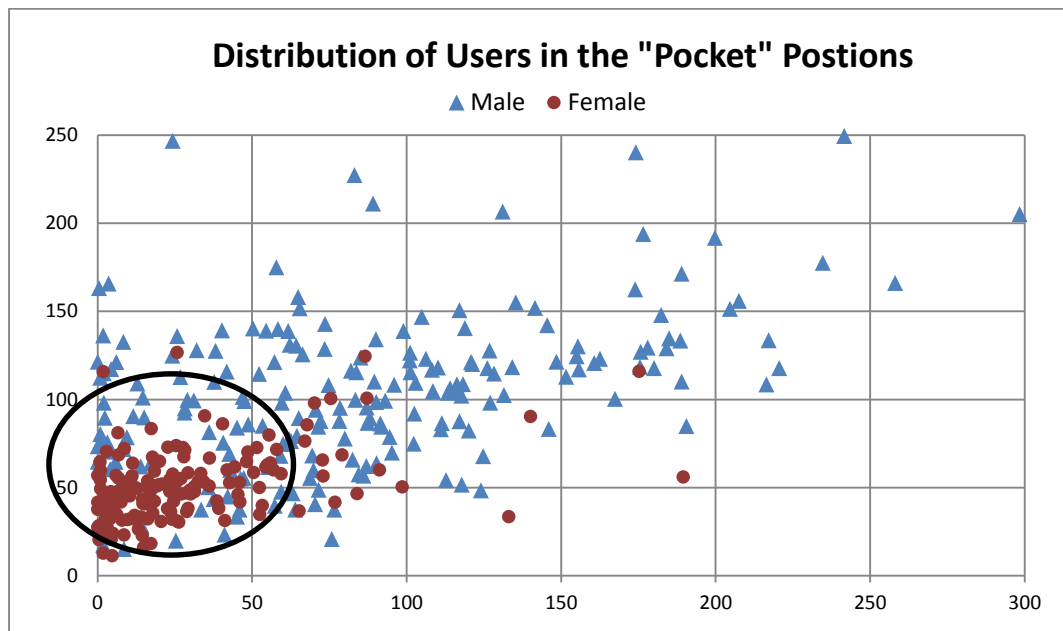
Once the information is processed and summarized correctly, machine learning algorithms are able to draw an imaginary circle and separate different groups, in this case men from women. Of course, we can see some users that do not behave according to their gender in this particular case. To improve the results we have built other methods and integrated them into a Meta algorithm that is able to synchronize them, predict based on what the majority dictates, and indicate a confidence level. This method is called *Voting*.

Most people are not aware, but accessing the installed application list does not require any permission from the user. Therefore, in order to create additional methods, we examined the types of applications that exist on the device. Android OS divides applications by type into 40 categories, such as: sports, games, navigation, music, etc. The average user is less concerned about sharing how many applications he or she has, in which category they belong, or their size (without supplying their specific names). And, although it may seem of little importance, we can infer quite valuable characteristics of the user only from the application description on Google Play. We managed to see that there are differences between men and women in this area too. Even if those apps change over time, machine learning algorithms are able to adapt.

**Gender Differences in App Categories**

Male ■ Female ■

*(Bar chart showing percentage split between Male and Female users across app categories: Music, Daily, Cards, Action, Race, Sports, Strategy, Adventure, Board Games. Music ~32% male, Daily ~30% male, Cards ~17% male, Action ~95% male, Race ~88% male, Sports ~92% male, Strategy ~82% male, Adventure ~81% male, Board Games ~78% male.)*

After testing these methods on actual people's devices, we discovered changes in patterns of use among men and women that were not previously known. We have seen, for example, that men tend to download more apps as well as relatively heavier apps (50Mb+), and non-free apps too—as is shown in the below graph:

**Differences in App types**

Male ■ Female ■

*(Bar chart comparing Male and Female for two app types: non-free and heavy. Non-free: Male ~30, Female ~7.5. Heavy: Male ~28, Female ~9.)*

By combining these algorithms and more, any app builder could infer the user's gender in a way that is fast and highly accurate without demanding permission to access the user's private data. The full method and results will be displayed in the annual conference for mobile research and development taking place in Italy this month (MobileSoft 2015).