Personalized Context-Aware Smartphone Usage Forecasting: A Machine Learning Model

Dr.M.Deepa, S.Nathiya, R.Ramya

Assistant Professor
Department of Computer Science,
Pavai Arts and Science College for Women
Namakkal, India

Abstract- In the past few decades, the field of human activity recognition has experienced significant growth. Numerous techniques for gathering data and evaluating it to find activity have been thoroughly researched. Contextual information pertinent to users' varied phone usage activities is captured through the device logs as a result of the growing popularity of recent enhanced features and context awareness in smart mobile phones. Context-aware customized systems may be created by simulating and forecasting how people will use their smartphones in different situations, such as temporal, geographical, or social information. In order to perform a context-aware analysis, we begin by applying five well-known machine learning methods for classifying data. We then present empirical assessments of an artificial neural network-based classification model, which is frequently used in deep learning, and perform a comparative analysis. A number of experiments are run on real mobile phone datasets gathered from users individually in order to evaluate the efficacy of these classifier-based context-aware models. Intuitive context-aware systems for smartphone users may be designed and built with the aid of the overall experimental findings and debates, which can be helpful to both researchers and application developers.

Index Terms- User behavior modeling, machine learning, Predictive analytics, and smartphone analytics.

I. INTRODUCTION

Smartphones are now regarded as the most practical and necessary items in our everyday lives, allowing people all over the world to interact with one another for a variety of purposes. In the modern world, 96.8% of people utilize mobile devices, and in many wealthy nations, this percentage even reaches 100% [1]. Data-driven personalized mobile services and systems are becoming an increasingly significant technology for creating and building user-centric smart mobile apps recently, thanks to the quick advancements in context-aware mobile technologies and the rising popularity of data science research [2].

In their varied day-to-day activities, people use smartphones for a variety of functions, including voice communication via phone calls, Internet browsing, utilizing mobile apps (Apps), e-mailing, online social networking, instant messaging, etc. [3]. Through the device logs, such as phone call logs [4, 5], app use logs [6, 7], notification logs [8] and web logs [9], individuals' such actions with their phones and associated contextual information, such as temporal, geographical, or social contexts, are recorded by their own devices. Classification is the most widely used machine learning approach to model and forecast individuals' such phone using behaviors in the field of mining mobile phone data.

Classification's objective is to correctly assign activity class labels to instances whose contextual characteristics or attribute values are known but whose class values are hypothetical [10]. It is difficult to effectively model and forecast user phone call activities from phone log data using machine learning approaches since different classifiers produce various predictions in various situations. As a result, we consider a thorough experimental examination of several machine learning technique-based models in this work that use contextual mobile phone data. We use the most widely used classifiers, such as Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), to assess the efficacy of a machine learning-based predictions model.

We pick these methods because they are well-known machine learning classifiers that are widely used to context-aware smartphone applications and systems for a variety of uses. For the benefit of end users of mobile phones, we perform our analysis by calculating the prediction results in terms of precision, recall, F-measure, kappa, receiver operating characteristic (ROC) value, and error rate measured by mean absolute error (MAE) and root mean squared error (RMSE) for all these classifier-based models.

The remainder of the paper is structured as follows. A summary of machine learning classification methods and related efforts in the field of context-aware mobile services and systems is given in section II. In part III, we briefly go over multidimensional contexts and the accompanying context-aware mobile services. We provide the methods for estimating customized smartphone usage in section IV. In section V, we provide the experimental findings and talk about the efficacy.
of the classification models. Finally, the "Conclusion and future work" section brings this research to a close and discusses the work to come.

II. MACHINE LEARNING CLASSIFIERS

One of the most used artificial intelligence techniques, classification uses a model learned from training data to predict the class of fresh samples. In general, classification is described as a learning technique that assigns or categorizes data examples to specified class labels in a dataset. Data classification, according to Han et al. [10], is a two-stage process. The first phase is learning, during which a model for categorizing data is created using a given dataset. The testing set is the set of data from which a classification function or model is learned. The testing set is the set of data used to evaluate the classification performance of the learned model or function. Naive Bayes Classifier

One of the most well-liked classification methods in the field of data mining is Naive Bayes (NB) [11]. A Naive Bayes classifier is a fundamental probabilistic-based method that can predict the probability of class membership [10]. It may easily accommodate the missing attribute values by simply removing the relevant probabilities for those characteristics when calculating the likelihood of membership for each class. Class conditional independence describes how, in a Naive Bayes classifier, an attribute's impact on a specific class is also independent of that of other attributes.

Decision Tree (DT) Classifier

Decision trees are a widely used and debated tool for classifying data before using it to make predictions [12]. Quinlan's ID3 is the fundamental decision tree construction method [12]. The ID3 algorithm builds a decision tree using a top-down method, testing each characteristic or context at each node using a greedy search over the provided training dataset. To determine which characteristic to test at each node in the tree, it estimates the entropy and information gain, two statistical properties [12]. Quinlan [13] has suggested the C4.5 algorithm, an extension of the ID3 method. Similar to ID3, C4.5 creates decision trees from a training dataset. This algorithm provides contextual decision rules to anticipate user behavior on mobile devices. Predictive modeling is carried out using the C5.0 modified decision tree method [10]. Decision tree C5.0 is substantially quicker and uses less memory than C4.5. Different methods are used by researchers to create decision trees for their study. Decision tree classifiers are mostly used to assess contextual mobile phone data in the context of context-aware mobile services and systems.

Random Forest (RF) Classifier

The Breiman et al.'s Random Forest Classifier [14] is an ensemble machine learning approach that combines many learning algorithms to provide a forecast. Using random feature selection and bootstrap aggregation (bagging), random forest creates a set of decision trees with controlled variance. In order to estimate the final output activity class of the mobile phone users using their phone log data, random forest constructs a number of decision trees as opposed to a single decision tree. It lessens the over-fitting issue that the single decision tree outlined previously creates by producing many decision trees for a given dataset.

In the field of context-aware mobile services, a lot of studies employ the random forest classifier. For instance, Pielot et al.'s [15] prediction model made use of the random forest classifier. Pielot et al. [18] employed this classifier in their examination of mobile services in a different research. In their context-aware mobile apps, Bedogni et al. [16] employ random forest classifier using transportation mode identification methods. This classifier is used by Bayat et al. [19] in their research on identifying human activity from smartphone accelerometer data. In their work on activity identification utilizing data from mobile phones, Ayu et al. [20] employ the random forest classification approach. Random forest classifier is another tool used by Turner et al. [17] in their research and prediction of interruption management.

Support Vector Machines (SVM) Classifier

SVMs (Support Vector Machines) [15] are yet another well-liked classification method that are frequently utilized for many types of predictive analytics. The information foci in the information variable space are often separated by their class, either class 0 or class 1, using SVM as a binary classifier. In order to do this, a vector machine hyper plane a line that can traverse the variable space is selected. The vector machine learning computation uses this hyper plane to obtain the coefficients resulting in the best class separation. For a variety of reasons, researchers in the field of context-aware mobile services employ SVM classifier.

For instance, Pielot et al. [18] have examined contextual mobile phone data using Support Vector Machines. In their context-aware mobile apps, Bedogni et al. [16] employ SVM classifier using transportation mode identification methods. In their work on identifying human activity using accelerometer data from cellphones, Bayat et al. [19] utilize SVM classifier. Ayu et al. apply the SVM classification approach in [20] for their study on activity recognition utilizing data from mobile phones. This classifier is used by Fetter et al. in [22] to analyze and forecast selective availability for instant messaging.

K-Nearest Neighbor (KNN) Classifier

K-nearest neighbors (KNN), a different classification method, is one of the simplest classification techniques in machine learning [22]. It is a form of lazy learning, or instance-based learning. The entire calculation is postponed until classification in this classification strategy, which takes into consideration local approximation. It maintains every
instance that is present in the provided dataset and categorizes additional cases using similarity metrics like Euclidean distance. The K most similar events, also known as the neighbors, are then managed by searching through the whole preparation set for a subsequent test information point. The anticipated result is then produced by abridging the yield variable for those K scenarios that depend on a majority vote from the neighbors casting a bigger share of the vote. KNN classifier is a tool that many academics in the field of context-aware mobile services employ.

**Artificial Neural Networks (ANN)**

Deep learning is a field of machine learning that typically employs an artificial neural network (ANN), a computer model that takes inspiration from the biological neural networks seen in the human brain. Back propagation [10], which executes learning on a multilayer feed-forward neural network, is the most widely used neural network method. The network learns throughout the learning phase by iteratively modifying the weights to anticipate the right class label of the input information.

Alawnah et al. [21] develop a method for estimating the power consumption of smartphones using neural networks and power-related data. In order to forecast the factors that would influence the acceptability of mobile credit cards for contactless payments in the future, Leong et al. [22] suggest a neural network technique. In order to anticipate the factors that will influence the adoption of mobile commerce—the purchasing and selling of products and services via wireless portable devices—Chong et al. [13] offer a neural network technique. Tan et al.'s [14]

### III. CONTEXT & CONTEXT AWARE MOBILE SERVICES

Any information that may be utilized to describe the condition of an entity is often referred to as context [13]. The individual smartphone user is considered an entity in this work. The important factors, such as temporal, geographical, or social information, which have an impact on persons for such mobile phone activities are needed to be taken into consideration as we intend to study individual's phone call activities utilizing machine learning classification algorithms. We briefly describe such contextual data in the context of context-aware mobile services and systems in the sections that follow. Which are:

**Temporal Context**

One of the main contexts influencing a user's phone call actions is the temporal context. For instance, using a cell phone in the morning can differ from doing so in the evening or at night [14]. Since every user action in the actual world is tied to a specific timestamp (e.g., YYYY-MM-DD hh:mm:ss), temporal context is crucial for simulating people's mobile phone usage patterns. The examination of a sizable sample of user data by Halvey et al. [12] has demonstrated that the day of the week plays a significant role in predicting mobile user behavior. The precise time of day (e.g., hh:mm:ss), the specific date (e.g., YYYY-MM-DD), and the days of the week (e.g., Monday, Tuesday, ...., Sunday) on which an action will take place are all examples of the temporal context. Additional temporal information, such as public holidays, weekdays, and weekends, may have an impact on decision-making for mobile phone users in addition to these key temporal details.

**Spatial Context**

Another important user context that may be utilized to more successfully model and anticipate a person's mobile phone usage is the user's physical location [21]. The rationale is that a person's phone call actions might be considered a location-based service. For instance, one's activities at work could not be the same as those she engages in at home. Thus, for context-aware apps to be able to offer location-based mobile services for people, understanding user mobility in their everyday lives is a vital challenge. Mobile phone applications that use the user's current location or spatial context are very popular for two reasons: (1) location-based mobile services depend on knowing the user's geographic availability to obtain pertinent information on the area or spot, and (2) the user behaves appropriately in that specific spatial context [20]. So, in our study, we consider the individual's position as a geographical context.

**Social Context**

In addition to the spatio-temporal factors mentioned above, social settings also affect how each individual mobile phone user makes decisions [20]. In the actual world, people participate in a variety of social events including lectures, seminars, and business meetings. According to Sarker et al. [22], people significantly differ from one another in how they use their mobile phones while participating in various events. For instance, one mobile phone user could be content to receive incoming calls during a business meeting, but another person would not want to do the same things at such time [15]. Even a specific person's behavior may vary depending on the sort of incident that occurred [13]. Even a specific person's behavior may vary depending on the sort of incident that occurred [13]. For instance, a person's phone call behavior or answer at a "professional meeting" can be very different from what she does in a "lunch-break" situation. Te social environment, such as the ties between people, e.g., family, friends, colleagues, love partners, or other relationships, have a great influence on individual mobile phone users to adopt such phone call handling decisions in such situations [21, 22]. For instance, when attending an important meeting, a person normally "declines" incoming calls; yet, if the contact is from her "boss," she "answers."
IV. EXPERIMENT RESULTS & DISCUSSIONS

We have carried out a variety of experiments on the real mobile phone datasets for forecasting individual mobile phone users' activity in various scenarios, in order to assess the performance of each machine learning classifier based model. Ten phone log datasets, each containing a person's varied phone call actions and associated contextual data, were used in our research.

Over the course of a year, these datasets were gathered from a variety of users, including students, academics, and staff. The datasets include people with various calling examples, call dissemination in multi-dimensional contexts, such as the temporal, spatial, and social contexts discussed above, and corresponding phone call activities, such as incoming call responses, which include users' answering or declining calls, missed calls, and outgoing calls of every person.

We estimate the precision of each classifier-based model in terms of Precision, recall/sensitivity, and ROC in order to assess the prediction accuracy by contrasting the expected response with the actual response (i.e., the ground truth). According to the CCI (correctly classified instances) rate, ICI (incorrectly classified instances) rate, and receiver operating characteristic (ROC) value of various individual mobile phone users using their own mobile phone datasets, we have shown the prediction results of each classic classifier-based model in Tables 1 below.

Table 1: Results of multiple machine learning classification models employing real-time mobile datasets for predictions.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>CCI (%)</th>
<th>ICI (%)</th>
<th>ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>82.53</td>
<td>19.74</td>
<td>0.93</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>95.93</td>
<td>13.86</td>
<td>0.96</td>
</tr>
<tr>
<td>Random Forest</td>
<td>95.36</td>
<td>12.62</td>
<td>0.96</td>
</tr>
<tr>
<td>SVM</td>
<td>92.61</td>
<td>17.38</td>
<td>0.89</td>
</tr>
<tr>
<td>KNN</td>
<td>71.48</td>
<td>18.51</td>
<td>0.73</td>
</tr>
</tbody>
</table>

For instance, Table 1 demonstrates that RF and DT examples are accurately identified at a rate of 95.93% and 95.36%, respectively, which is greater than that of other classifier-based models.

Figure 1: Effectiveness Comparison Result for various machine Learning Classifiers

CONCLUSION & FUTURE WORK

Context-aware, adapted smartphone usage modeling and prediction may be used to create a range of data-driven, smart apps that can help end users of mobile devices intelligently. We have examined the efficacy of many prominent machine learning classifier based use models for forecasting, including neural network learning model using real-world mobile phone data, for intelligently delivering such context-aware tailored services. We include multi-dimensional circumstances, such as temporal, geographical, or social contexts, that have an impact on people's phone use, either independently or in combination, in our machine learning-based effectiveness study.

Despite using phone calls as a scenario, the results of this study may be applied to various mobile application domains based on pertinent multi-dimensional settings and users' smartphone usage. We think that by providing academics as well as application programmers with a reference point for their work on context-aware mobile services and systems, our efficacy study on various machine learning classification-based context-aware models might be helpful. The development of a context-aware real-world application using user data from smartphones, in order to offer intelligent tailored services...
to the end users of mobile phones, and to evaluate our work at the application level, might be future work based on our effectiveness study.

REFERENCES: