Towards Intelligent Decision Making in Emotion-aware Applications

Jiaqi Wang\textsuperscript{1,2,3}, Siqi Wang\textsuperscript{4,1,2}, Yanxiang Guo\textsuperscript{1,2,3}, Xiping Hu\textsuperscript{1,2}, Xitong Li\textsuperscript{5}, Jun Cheng\textsuperscript{1,2}

\textsuperscript{1}Shenzhen Institute of Advanced Technology (SIAT), Chinese Academy of Science, Shenzhen, China
\textsuperscript{2}The University of Hong Kong, Hong Kong, China
\textsuperscript{3}The University of Hong Kong, Hong Kong, China
\textsuperscript{4}National University of Defense Technology, Changsha, China
\textsuperscript{5}HEC, France

wjq1015@hku.hk, wangsiqui0c@gmail.com, yxguo@hku.hk, xp.hu@siat.ac.cn, lix@hec.fr, jun.cheng@siat.ac.cn

Abstract

In this paper, we propose an intelligent emotion-aware system (IES), which aims to provide a systematic approach that can make use of the online technology to improve the intelligence of different emotion-aware mobile applications. IES is constructed to provide multi-dimensional online social community data collection and processing approaches for decision making, so as to recommend intelligent services for emotion-aware mobile applications. Furthermore, we present a flow of intelligent decision making process designed on IES, and highlight the implementation and orchestration of several key technologies and schemes applied in this system for different emotion-aware mobile applications in run-time. We demonstrate the feasibility of the proposed IES by presenting a novel emotion-aware mobile application - iSmile, and evaluate the system performance based on this application.

1. Introduction

In this day and age, more and more emerging mobile applications and services can be foreseen and applied in human society. In the meantime, emotion-aware computing (EMC) is playing an important role in various realms, such as healthcare research, disaster relief and emergency response \cite{1}. Therefore, there is a natural connection between EMC and mobile applications, which gives rise to emotion-aware mobile applications \cite{2}. For emotion-aware mobile applications, it is crucial to realize an intelligent decision making process, so as to effectively identify the mental health states of diverse users in time and precisely determine which tips or customized services to be recommended to individual user \cite{3}.

However, existing emotion-aware mobile applications generally rely on limited small-scale data collected by individual smartphone for decision making, leading to an insufficient accuracy of emotion recognition \cite{4}. More importantly, as the technology advances, the capabilities of mobile devices have been significantly enhanced by stronger computational resources, multiple radios and a variety of sensors, which enables mobile devices to collect much richer context information from the user’s daily life than before. However, such rich context information, like user environment and user emotion state history, is still neglected during the decision making process of emotion-aware applications. Besides, the traditional approaches for emotion-awareness are based on large images/video analytics that require high bandwidth multimedia communication support, which may not always be available in mobile situations \cite{5}.

Motivated by those gaps above, this paper proposes an intelligent emotion-aware system (IES) to achieve intelligent decision making and content recommendation for emotion-aware mobile applications by collecting and processing multidimensional online social community data. Our contributions in this paper can be summarized as follows:

- We explore a seamless systematic solution that utilizes online technology to enable intelligent decision making in emotion-aware mobile applications.
- We provide multi-dimensional emotional data collection and processing approaches that enable efficient and effective emotion-aware computing in mobile applications.
- We demonstrate the feasibility of the proposed IES by a novel emotion-aware mobile application called iSmile, which is designed to recommend proper tips for healthy and good sleep.
Experiments show that our system can increase positive emotion by 14.57%.

The rest of this paper is organized as follows. Section 2 introduces the research background of our works. Section 3 presents the system design for smart alarm sound recommendation system named iSmile, and section 4 describes the implementation strategies from the perspectives of the mobile side and the cloud side. Section 5 demonstrates the features of our works through application cases. Section 6 shows a case study of our system and evaluates the performance through practical experiments. Finally, section 7 concludes this paper.

2. Research Background

Emotion-aware systems that use context information for intelligent decision making have attracted attentions for years. For example, Big Data Application in Emotion-aware Healthcare (BDAEH), a system that can collect, transmit, store, and analyze healthcare data for emotion computing, is used to improve healthcare services by providing users or medical staff with analytical results from the emotion-aware decision making [6]. Hossain and Muhammad [7] propose an emotion-aware healthcare system using speech and image signals to detect users’ emotions before classification to determine whether healthcare services are needed. Health-CPS [8] concentrates on supporting convenient services and environment for patient-centric healthcare applications through cloud and big data for intelligent decision making.

Decision making, as another popular research area, plays an important role in recommendation systems and online social networks. Giacchi et al. [9] model the dynamic behaviors of nodes inside a social network to comprehend the decision-making process in social networks, by which predicting and analyzing new behavioral patterns can be predicted and analyzed. A real-time web mining application focusing on the trend analysis to support efficient and effective decision making is presented in [10], which is achieved by analyzing various data from the selected web sites. Such context mining on online social networks for decision making can be easily extended to diverse application fields, such as financial market decisions, healthcare systems, and emotion-aware systems.

In our previous work, the framework of Emotion-Aware Smart Tips (EAST) has been implemented to recommend smart tips to users for healthier sleep and emotional states by collecting acceleration data from users’ mobile phones [11]. Sleep patterns are first extracted, and then EAST uses the multi-variate neural random forest model to determine the relationship between sleep patterns and emotional states. In this work, based on the work of EAST, we further propose IES as a seamless systematic model to provide intelligent decision making for emotion-aware mobile applications by leveraging several online technologies. Making use of multi-dimensional emotional data (e.g., from online social networks) and processing methods for decision making provided by IES, iSmile is able to recommend a variety of intelligent services and practically improve the sleep quality of users.

3. System Design

As an instantiation of the constructed IES, we propose a novel online emotion-aware mobile application named iSmile. iSmile can exploit multi-dimensional online social community data to enable smart decision making on alarm sound recommendation, so as to improve sleep quality of users. The system design of iSmile is shown in Figure 1. There are five key components in iSmile: Raw Data Collection (RDC), Feature Extraction, Data Integration, Smart Alarm Recommendation (SAR) and Feedback Calculation. Figure 2 is the workflow of our system. After a user sets the alarm clock, RDC starts working on collecting raw feature data defined and used by our system, for example, raw acceleration data for user’s sleep status. At a predefined interval (e.g., 2 minutes) before alarm time, raw data collected by RDC are sent to Feature Extraction to extract Feature Vector. Data Integration integrates Feature Vector extracted by Feature Extraction with context-aware information into our system as User Context Model (UCM). SAR is the
center component for smart decision making on alarm sound recommendation. SAR takes UCM as input, with the help of big history data in User History Record (UHR), recommends proper alarm sounds for users. We describe the details of each component as follows:

### 3.1. User Context Model

UCM is designed to represent different users in different situations. It consists of three parts: Feature Vector, Context Vector and Social Vector. These parts have different functions:

**Feature Vector:** This vector is used to include specific features, which are needed by smart alarm sound recommendation. For example, if the goal is to recommend smart alarm sounds that wakes users up in the morning, sleep features can be used as the feature vector.

**Context Vector:** It takes context information into account, including user’s emotional states and real-time weather. We use arousal-valence model [12] to represent user’s emotion states.

**Social Vector:** People with different social information may have different tastes on smart alarm sounds in a certain situation. Thus, iSmile exploits social information in terms of five aspects: Gender, Nationality, Age, Occupation and Academic Degree. All above social information can be provided by an online social application.

Multi-dimensional data from these three parts above can be obtained from multi-source sensors and social network applications on smartphones. They are then fed into SAR to make smart decision and recommend proper alarm sounds.

### 3.2. Smart Alarm Sound Analysis

Before making a decision on recommending alarm sounds, we first need to analyze alarm sounds to obtain arousal valence emotion states expressed by each alarm sound. Based on the method mentioned in [13], we extract 6 acoustic features from each alarm sound: zero-crossing rate (ZCR), tonal type (TT), tempo (TP), low energy rate (LER), spectral centroid (SC) and unit power (UP). These six features are integrated into an alarm sound feature vector \( AFV = (ZCR, TT, TP, LER, SC, UP) \). We define the Similarity (Sim) between alarm sound \( i \) and alarm sound \( j \) are defined as follows:

\[
\text{Sim}(AFV_i, AFV_j) = \frac{1}{6} \left( (1 - \frac{|TT_i - TT_j|}{\text{max}(TT_i, TT_j)}) + (1 - \frac{|TP_i - TP_j|}{\text{max}(TP_i, TP_j)}) + (1 - \frac{|ZCR_i - ZCR_j|}{\text{max}(ZCR_i, ZCR_j)}) + (1 - \frac{|LER_i - LER_j|}{\text{max}(LER_i, LER_j)}) + (1 - \frac{|SC_i - SC_j|}{\text{max}(SC_i, SC_j)}) + (1 - \frac{|UP_i - UP_j|}{\text{max}(UP_i, UP_j)}) \right)
\]

Our target is to assign each alarm sound a corresponding arousal-valence emotional state that is the most likely to be expressed by this alarm sound. We manually choose four alarm sounds, which are predominant in one dimension of the 2-dimension arousal-valence model, and manually label these four alarm sounds by subjective judgment. Then we adopt Algorithm 1 to analyze arousal-valence value of each alarm sound in our alarm sound library. When a new alarm sound comes to our system, its six acoustic features are extracted and Algorithm 1 is applied to these 6-dimension alarm sound features to obtain an arousal-valence value for the target alarm sound.

### 3.3. Smart Decision Making

With UCM defined in Section 3.1, we can calculate similarity between users based on their social background and context information. Then decisions on which alarm sound they would like to hear are made based on these similarities and UHR. Finally, the alarm sound in our alarm sound library, which is the closest to the prediction, is delivered to the mobile side. Specifically, for UCM that can be regarded as a vector \( (FV, CV, SV) \) containing different kinds values, the similarity between \( UCM_i \) and \( UCM_j \) can be computed
Algorithm 1 Alarm sound arousal-valence calculation.
\textbf{Require:} acoustic feature vector of input alarm sound
\textbf{AFV}, alarm sound library \textit{L}, \textit{K} most similar alarm
sounds to be considered

\textbf{Ensure:} (\textit{Arousal}, \textit{Valence}) for input \textit{AFV}
1: \textbf{initialize} an empty array \textit{Arr}, \textit{W}_{all} = 0
2: \textbf{for} every alarm sound recorded \textit{L}_\textit{i} in \textit{L} \textbf{do}
3: \textbf{calculate} \textit{S}_\textit{i} = \textit{Sim}(\textit{L}_\textit{i}, \textit{AFV}, \textit{AFV})
4: \textbf{push} \textit{(S}_\textit{i}, \textit{L}_\textit{i}, \textit{Arousal}, \textit{L}_\textit{i}, \textit{Valence}) into \textit{Arr}
5: \textbf{end for}
6: \textbf{sort} \textit{Arr} in descending order
7: \textbf{initialize} \textit{W}_{all} = 0
8: \textbf{for} each \textit{Arr}_\textit{i} from the top \textit{K} of \textit{Arr} \textbf{do}
9: \textit{W}_{all} = \textit{W}_{all} + \textit{Arr}_\textit{i}.\textit{Sim}
10: \textbf{end for}
11: \textbf{initialize} \textit{Arousal} = \textit{Valence} = 0
12: \textbf{for} each \textit{Arr}_\textit{i} from the top \textit{K} of \textit{Arr} \textbf{do}
13: \textit{W} = \textit{Arr}_\textit{i}.\textit{Sim} / \textit{W}_{all}
14: \textit{Arousal} = \textit{Arousal} + \textit{W} \times \textit{Arr}_\textit{i}.\textit{Arousal}
15: \textit{Valence} = \textit{Valence} + \textit{W} \times \textit{Arr}_\textit{i}.\textit{Valence}
16: \textbf{end for}
17: \textbf{return} (\textit{Arousal}, \textit{Valence})

by:
\begin{equation}
\textit{Sim} (\textit{UCM}_\textit{i}, \textit{UCM}_\textit{j}) = \textit{W}_\textit{f} \times \textit{Sim} (\textit{FV}_, \textit{FV}_j) + \\
\textit{W}_\textit{c} \times \textit{Sim} (\textit{CV}_, \textit{CV}_j) + \textit{W}_\textit{s} \times \textit{Sim} (\textit{SV}_, \textit{SV}_j)
\end{equation}

(2)

where \textit{Sim} (\textit{FV}_, \textit{FV}_j) and \textit{Sim} (\textit{CV}_, \textit{CV}_j) are defined as
the Euclidean distance of two vectors (each dimension of which is normalized as a decimal
between 0 and 1). It is not straightforward to deal
with \textit{Sim} (\textit{SV}_, \textit{SV}_j) since Nationality and Occupation
data in \textit{SV} are text values, so we implement an
ontology-based method to calculate similarity between
two English words by considering their semantic
relatedness [14]. Thus, similarity between \textit{SV} can be
calculated by:
\begin{equation}
\textit{Sim} (\textit{SV}_, \textit{SV}_j) = \textit{W}_{age} \times (1 - \frac{|\textit{age}_i - \textit{age}_j|}{\text{max} (\textit{age}_i, \textit{age}_j)}) \\
+ \textit{W}_{\text{degree}} \times (1 - \frac{|\text{degree}_i - \text{degree}_j|}{\text{max} (\text{degree}_i, \text{degree}_j)}) \\
+ \textit{W}_{\text{nationality}} \times \textit{Sim} (\text{Nationality}_i, \text{Nationality}_j) \\
+ \textit{W}_{\text{gender}} \times (1 - (\text{gender}_i \oplus \text{gender}_j)) \\
+ \textit{W}_{\text{occupation}} \times \textit{Sim} (\text{Occupation}_i, \text{Occupation}_j).
\end{equation}

(3)

Next, to recommend healthy alarm sounds to users,
we need a UHR. UHR is a table contains history user
records (i.e. \textit{UCM} and the preferred alarm sound for
corresponding user), which can be a useful reference
for smart alarm recommendation. After obtaining the

Algorithm 2 Desired alarm sound emotional states calculation.
\textbf{Require:} \textit{UCM}, \textit{UHR}, \textit{K} similar records to be
considered

\textbf{Ensure:} Desired (\textit{Arousal}, \textit{Valence}) for input \textit{UCM}
1: \textbf{initialize} an empty max priority queue \textit{Q}
2: \textbf{for} each \textit{UHR}_\textit{i} in \textit{UHR} \textbf{do}
3: \textbf{calculate} \textit{Sim}(\textit{UHR}_\textit{i}, \textit{UCM}, \textit{UCM})
4: \textbf{push} \textit{(Sim(\textit{UHR}_\textit{i}, \textit{UCM}, \textit{UCM}, \textit{Arousal}, \textit{UHR}_\textit{i}, \textit{Valence}))}
5: \textbf{end for}
6: \textbf{initialize} an empty array \textit{Arr}, \textit{Simsum} = 0
7: \textbf{for} \textit{i} = 0 to \textit{K} \textbf{do}
8: \textbf{current} = \textit{Q}.\textit{top}
9: \textit{Simsum} = \textit{Simsum} + \textit{current}.\textit{Sim}
10: \textbf{push} \textit{(Simsum, current.\textit{Arousal}, current.\textit{Valence})}
11: \textbf{end for}
12: \textbf{return} \textit{(Arousal}, \textit{Valence})

desired alarm sound emotional states which corresponds
to user in current situation, we can find the closest alarm
sound in a library.

3.4. UHR Update

As seen from above, UHR is vital important of
smart alarm sound recommendation. The more records
in UHR, the better the performance of SAR will be.
There are two ways to increase the number of records
in UHR: 1) manually insert; 2) insert with a feedback
calculation. For the first method, we can set up a
survey for potential users and insert results into UHR.
For the second method, our system can be started with
a relatively poor UHR (even a randomized UHR). During
the run-time of the system, we can use emotional states
in \textit{UCM} and in feedback to decide whether to insert this
record into UHR.
4. Implementation Strategies

4.1. The Mobile Side

To integrate our system on Android smartphones, a mobile application for Android is needed. There are several important aspects on mobile side:

- **Data collection**: IES makes use of multiple sensors on Android to collect sensor data needed for recommendation.

- **Communications**: With collected data, the mobile side application needs to communicate with the cloud side, so as to transfer sensor data and other required user information(e.g. social information) to the cloud side and retrieve the recommendation results. To achieve this objective, HTTP is used as the communication protocol between the mobile side and the cloud side.

Many different methods like [15] have been proposed to help design privacy-preserving systems. We preserve user's privacy in terms of three aspects: anonymization, secure transportation and data perturbation.

4.2. The Cloud Side

The backend of iSmile is built in the cloud. It is designed to support the mobile app by running a bunch of services. By various service APIs that cloud vendors provide, the cloud resources can be easily and rapidly integrated into our prototype system. From the functionality prospective, the backend is implemented as a bunch of RESTful web services. There are five main services:

- **FileUpload Service**: A POST method, which receives the raw data file and stores it in the cloud object storage. After that, it keeps the experiment ID and the object URI into the database, and then returns the object URI back as a response.

- **FeatureExtraction Service**: A GET method with an object URI as the parameter. Feature Extraction is then started to work on data pointed by URI to extract required features.

- **EmotionalStates Service**: A GET method with features extracted by FeatureExtraction Service, emotional states can be then calculated through those features. The calculated emotional states will be returned as the response.

- **AlarmRecommend Service**: A POST method that accepts the mood vector and the weather info and returns a music ID as the recommendation for the alarm sound.

- **Feedback Service**: A POST method that saves the experimenters’ feedback to the database and triggers UHR update mentioned in Section 3.2 if feedback calculation is deployed.

5. Application Examples

iSmile is enhanced by emotion-aware cognitive system [1] to demonstrate the system. The main function of iSmile is to detect users’ sleep status and collect the data, so that the backend can analyze them and push proper alarm sounds to users. We specifically design a refreshing and brief user interface of iSmile as the vision effect of an app has an important influence on users’ mood, which is exactly our concern.

Figure 3 shows some of the UIs of iSmile application.

Users will see this page after opening the app. They can set the time when they want to get up. There are tabs on the bottom of the screen. Users can click the right one to switch to the settings, where users can set whether they would like to choose the alarm sound by themselves or accept the system push. Once the users finish the setting, they can click “start” then enter the time-keeping page. User needs to guarantee the mobile phone has enough power and good network. When time is up, the alarm sounds and users need to feed back their moods after they close the alarm. As shown in Figure 3(b), (c), (d), different alarm sound pages are also designed to help users feel well.

Polysomnography (PSG) is an important tool to monitor sleep quality [16]. However, special devices and professional knowledge are required in PSG-based tests, making it hard to be implemented in a normal bedroom. We use 3-dimension acceleration on mobile phones to collect acceleration data at a frequency of 5 Hz. The acceleration data will be uploaded to server for sleep feature extraction. The location information will also be uploaded. On the cloud side, we apply feature extraction and mood prediction mentioned in [11]. The acceleration data are divided into windows, which contains 20 frames (i.e. 10s). Then, statistical values mentioned in Section 2 are calculated. User emotional states are also predicted with these statistical values. After calculating user’s preferred alarm sound
with algorithms mentioned in Section 3, the server will inform Android client which alarm sound to be played. After closing the alarm sound, users will enter the feedback page. They need to drag the progress bars to choose how they feel and what their moods are. Then once users click “OK”, clients will submit two features, arousal and valence, to the cloud side with Feedback Service.

6. Initial Experiments

To evaluate the effectiveness and efficiency of our system, we design and conduct experiments with iSmile application. In the rest of this section, we discuss the overhead of iSmile mobile application on a popular smartphone Meizu Note 5.

6.1. System Overhead

We evaluated the basic overhead of iSmile running on mobile phone in four aspects: (i) CPU, (ii) RAM, (iii) Battery, (iv) Network. We conduct the evaluation using Meizu Note 5 to run 10 nights where every time lasts for 10 hours. The results are shown in Figure 4 and Figure 5.

**CPU:** The CPU, which is HELIO P10 in Meizu Note 5, usage is around 2% during running, as shown in Figure 4(a). The main CPU occupation operation is collecting acceleration data, which is uniform in the whole procedure, so there is not much pressure on CPU. The CPU usage value is of the whole system including other apps and system occupation. Therefore, iSmile performs well in CPU usage.

**RAM:** Figure 4(b) shows that RAM usage of iSmile is also inconspicuous, where the whole RAM usage is around 50MB. The data are stored in RAM until wrote into flash as file before uploading. The size of file is around 5MB, which depends on the length of test time.

**Network:** Network related operations are not frequent, which is mainly before the alarm sounds when iSmile uploads the data. Network consumption is around 5MB every night, slightly larger than the data file size.

**Battery:** We test 10 hours as a standard test time, with other apps killed, and the battery consumption is 26% on average (the battery capacity in Meizu Note 5 is 4000mAh). The power is consumed more when iSmile uploads data to server but the consumption curve is relatively uniform as the whole.

6.2. Impact of iSmile

**Experimental Settings:** We involve eight subjects (three females and five males) into our experiment lasting for ten days. Subjects are asked to start iSmile and set the alarm clock before they go to sleep. iSmile records their sleep data and uploaded data to the cloud side for alarm sound recommendation. To minimize the differences caused by different smartphones, we install iSmile on several Meizu Note 5 smartphones with the same setting (e.g. volume) and deliver these prepared smartphones to subjects. Some changes are made to iSmile for experiments. We add a trigger in the cloud side, which can control the cloud side to push either a smart alarm sound recommended by our system or a randomly chosen alarm sound from the alarm sound library consists of nine alarm sounds in different styles. In this way, subjects do not need to manually change the working mode of iSmile on their smartphones, so the principle of control experiments will not be violated. We randomly chose five of the ten days to open the trigger (i.e. smart alarm sounds recommended by our system are delivered to the mobile side), and leave the trigger closed in the rest five days (i.e. push randomized alarm sounds to the mobile side). Besides, the serial number of a smartphone is uploaded to the cloud side in the specified version of iSmile modified for this experiment, because we need to track each single user in this experiment.
Experimental Results: In total, data of 69 nights are recorded during the experiment, including 37 nights with smart alarm sounds recommended by iSmile (i.e. the treatment group) and the rest 32 nights with alarm sounds randomly chosen from the alarm sound library (i.e. the control group). The arousal value and valence value are normalized in the range of 0 to 10, and a higher value indicate a better emotional states.

Figure 6 illustrates the comparison results of arousal values between the treatment group and the control group. The average of arousal values of the treatment group is 6.41, while the average arousal of the control group is 5.81. The comparison results of valence values between these two groups are shown in Figure 7. The average value of valence of the treatment group is 6.76, while the average valence value of the control group is 6.13. The average arousal value and valence value of the treatment group are increased by 10.33% and 10.27% respectively, which indicates that iSmile can improve subjects’ emotional states compared to the random alarm sounds. Little difference is observed between males and females in our experiment. Considering that an emotional state can be viewed as a point in Arousal-Valence plane, we can calculate the positive emotion improvement by \( \sqrt{10.33\% \times 10.33\% + 10.27\% \times 10.27\%} = 14.57\% \).

7. Conclusions

In this paper, we have presented a novel IES, which is instantiated by a context-aware smart alarm sound recommendation system named iSmile, to enable a smart decision making process of emotion-aware mobile applications for customized service recommendation. IES can also be implemented to serve people in other areas. For example, smart reminders can be developed with IES to provide smart reminders for citizens to make the city life smarter. Students can also benefit by developing a system based on IES to give smart recommendation for online education. Our IES takes advantage of the sensing and computing capacity of smartphones and mobile crowdsourcing to collect multiple sources of online sensing data, so as to construct the UCM for smart alarm sound recommendations. More importantly, this system not only considers the context information, of alarm sounds, but also the comprehensive user context information including emotions and social information to intelligently decide preferable alarm sounds recommended to users in a variety of situations. Our practical experiments demonstrate that the proposed IES works effectively and efficiently in smart alarm sound recommendation tasks, and improves their emotional states by about 14.57% in contrast to traditional alarm sound delivery when waking people up in the morning. The success of iSmile demonstrates the effectiveness of the proposed IES and reveals its great potential in facilitating the intelligent decision making of emotion-aware mobile applications.

8. Acknowledgments

This work was partially supported by the Shenzhen-Hong Kong Innovative Project under Grant SGLH20161212140718841, the National Basic Research Program of China (973 Program)(no.2014CB744600), the National Nature Science Foundation of China (61632014, 61210010, 61772508, 81401570)

References


Figure 6. Comparison results of arousal values between the treatment group and the control group. (The higher, the better.)

Figure 7. Comparison results of arousal values between the treatment group and the control group. (The higher, the better.)


