

User Profile System Based on Sentiment Analysis for Mobile Edge Computing

Sang-Min Park¹ and Young-Gab Kim^{2,*}

Abstract: Emotions of users do not converge in a single application but are scattered across diverse applications. Mobile devices are the closest media for handling user data and these devices have the advantage of integrating private user information and emotions spread over different applications. In this paper, we first analyze user profile on a mobile device by describing the problem of the user sentiment profile system in terms of data granularity, media diversity, and server-side solution. Fine-grained data requires additional data and structural analysis in mobile devices. Media diversity requires standard parameters to integrate user data from various applications. A server-side solution presents a potential risk when handling individual privacy information. Therefore, in order to overcome these problems, we propose a general-purposed user profile system based on sentiment analysis that extracts individual emotional preferences by comparing the difference between public and individual data based on particular features. The proposed system is built based on a sentiment hierarchy, which is created by using unstructured data on mobile devices. It can compensate for the concentration of single media, and analyze individual private data without the invasion of privacy on mobile devices.

Keywords: User profile, sentiment analysis, mobile edge computing, social network.

1 Introduction

As information and communication technology has improved, the type of user data is moved from informative to sentiment data. That is, the user has started sharing their thoughts and feelings in the form of messages through social networking. In addition, as the amount of social network service (SNS) data has increased, user-generated data has increased in significance and has become useful for user preference analysis. Moreover, the refined user preference can provide more affluent and personalized data to application. For example, some commercial web site provides customer's age and location to complement customer's comments as background information. In addition, the refined user preferences which provides on proposed system can make better understanding for the review of purchased customer. However, to extract meaningful value from the SNS data, that is, to make the refined user preferences for a specific user, sentiment analysis, which is a useful

¹ Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea.

² Sejong University, 209, Neungdong-ro, Gwangjin-gu, Seoul, 05006, Korea.

* Corresponding Author: Young-Gab Kim. Email: alwaysgabi@sejong.ac.kr.

method that determines an individual's preferences and constructs a profile through accumulated personal sentiment analyses, is required for handling emotional data of the user, because SNS data include a large amount of personal feelings as compared with facts.

The user profile is a way of describing user characteristics using collected information about what the user prefers. In the past, user profiles were mainly used for the purpose of personal recommendation and personal search, which is a customized result based on user preferences for its own purpose [Sugiyama, Hatano and Yoshikawa (2004)]. For this reason, the user profile generated by a specific user cannot be used for other applications with a general purpose. Therefore, each application has to construct their own user profile with limited data of user behaviors, which is gathered from their application usage of users, without using personal data (e.g., user's interest etc.) that was otherwise accumulated on the mobile device.

That is, for the aspect of configuration, existing user profiles are generated using user input, tagging information, and the relationship with other users [Cai and Li (2010)]. Furthermore, due to the lack of user input data and the privacy concerns surrounding user data, it is difficult to obtain sufficient user data for a structural analysis. In addition, server-based approaches, which transmit user data (e.g., user preference, behaviors, age) from a client (e.g., diverse user's mobile device) to the server and construct user profile on the server, have a risk of diverse security attacks.

Therefore, in order to overcome these limitations, we proposed a sentiment user profile analysis, which can analyze individual user characteristics based on the forgetting curve. However, the past proposal [Park, Baik and Kim (2016)] only focus on applying the forgetting curve of Ebbinghaus with the weight on time to calculate the sentimental difference. In this paper, we expanded our past proposals [Park, Baik and Kim (2016); Park, Kim and Baik (2016)], describing more in detail the problem statements (i.e., in an aspect of data granularity, media diversity, and client-server solution), data acquisition type (i.e., application with internal storage, external storage, and without storage), and user-centric profile utilization (i.e., profile storage, individual polarity, and porting with polarity). The proposed system uses the user's private data (e.g., from chats, messages) against external public data (e.g., reviews and blogs). Furthermore, in order to integrate the data on various applications, which is scattered on diverse mobile devices, the system utilizes temporal clustering with the weight of the Ebbinghaus forget curve.

The main contributions of this paper are as follows: 1) To compensate the concentration of single media, the proposed system construct user profile with a sentiment hierarchy by using unstructured mobile data. That is, the proposed method manages sentiment user profile for each topics and can be used to sharing refined user preferences to new application. 2) It can reduce human labor inputs to provide user preference on each application 3) Furthermore, it supports privacy preserving analysis with individual private data on mobile devices. 4) In addition, the proposed system can reduce sharing user private raw data for analyzing user preferences based on mobile edge computing, which can distribute processing and resources from a server to a mobile device at the edge of the network. It can also be used to manage sharing their preferences under control of user.

The remainder of our paper is organized as follows. Section 2 describes the problems relating to data shortage of fine-grained analysis, the fragmentation of single media, and

the privacy issue of the server-side analysis for sentiment user profiling in mobile devices. Section 3 surveys recent research and technologies for a sentiment analysis and temporal analysis. In Section 4, we explain the equations used for calculating individual private data compared with external public data. In Section 5, we describe the architectural design of the proposed system as well as its implementation. We also demonstrate the feasibility of the system through several tests for user characteristics extraction in Section 6. Finally, Section 7 provides some concluding remarks and areas of future research.

2 Problem statement

In this section, we discuss the problem with the existing mobile user profile system in terms of data granularity, media diversity, and client-server solution.

In regard to the aspect of data granularity for analyzing user profiles, Whitelaw et al. [Whitelaw, Garg and Argamon (2005); Stoyanov and Cardie (2008)] have proposed a fine-grained sentiment analysis considering the granularity of data. The subject of product analysis has changed from analyzing the overall product to analyzing its fine-grained features (e.g., from the mobile phone to the battery) [Liu (2010)]. For example, sentiment analysis previously focused on whether the user has a positive or a negative impression when using the movie review data. As the objective of the analysis is more specific, features such as actors and effects become the subject of the analysis. This increases the preference detail such as *'the actor's performance was great but the film left an immature impression'*. However, to perform the analysis on such a fine-grained feature level, a relatively large amount of data is required compared with a coarse-grained analysis because the data classification at this feature level is detailed. Furthermore, a fine-grained analysis of a feature-level has the possibility to generate insufficient results, a situation which is referred to as a cold start, because it is difficult to obtain sufficient data compared with a coarse-grained analysis.

From a media perspective, a user profile analysis based on single media presents two forms of risk: concentration and specialization. A single media-based approach is used to analyze previous records for the media or similar users in the same media. However, a single media analysis system for general purposes has a problem with the preponderance of the user group and the specialization in a particular purpose. Analysis based on a single media can be concentrated on the user's age or groups according to the characteristics of the media. That is, when a particular subject is interested in the member of the media, appropriate results can be analyzed, otherwise that media is not suitable for analysis according to the characteristics of the topic. Moreover, users tend to prefer a specialized social service for each purpose, rather than a variety of functions that are supported by one particular medium. For example, people use a single medium such as Facebook for uploading pictures and sharing their thoughts. However, media trends which focus on their own purpose cause user data to spread to various media such as Instagram for pictures, Facebook for friends, and Twitter for the public. In addition, in respect to the acquisition of information for analysis, users acquire the information using more than one particular medium. For example, general information can be obtained through an Internet search, and the opinions of friends are mostly gained using a messenger service or SNS. That is, a conventional single media-specific analysis cannot be considered to be a comprehensive user analysis because media

have become more diverse. As a result, due to the preponderance of user groups and the specialization in a particular purpose, it is necessary to perform a user-centric analysis based on the users with respect to a variety of media.

User-centric analysis integrates various user information as single path to general understanding for a specific user. An integrated media approach used a hub for displaying a variety of media in a unified environment, such as rich communication suit (RCS) and Social-On. However, the use of an artificially integrated service is not suitable for the development of applications through an open platform and open market. User information is scattered across applications, so a common parameter is needed to integrate in single path. Temporal information can be one of solution because it used as a common parameter for various application. When temporal distance between occurrences is small, the temporal similarity of a topic is high rather than long temporal distance. However, it is difficult to obtain uniform temporal information because each medium uses its own different time unit such as day, hour, and second. Moreover, temporal information generated by a mobile device requires less modification because the media operations are based on a common clock.

Finally, the existing mobile profiling system is classified as being either a server-side or client-side approach. Most of the media collect user information such as writings and images on the client application and this information is then sent to the server-side to analyze user behavior. As the importance of privacy have been enhanced, server-side solution is difficult to analyze individual private data. Server-side approaches require that the user data stored in mobile devices have to be transmitted to the server and analyzed in the server [Brusilovskv, Stock and Strapparava (2000)]. However, as mentioned previously, privacy issues can arise when transmitting user data [Sateli, Cook and Witte (2013)]. To protect personal information, some studies have removed individual recognizable factors or have encrypted the data using various cryptography techniques. Deletion of the personal recognizable factors makes it impossible to perform a personalized analysis of a specific user [Liu, Jiang, Sha et al. (2012)]. Furthermore, although cryptography techniques can be used to protect user data from an interception attack, transferring a user's personal information to a server exposes the information to potential hacking. Some studies have collected information by using the user email address to crawl multiple SNS, in which personal information is maintained on the Web [Bird, Gourley, Devanbu et al. (2006)]. Unique parameters for media integration are used based on the user's email address or phone number for personal recognition. However, users sometimes use a plurality of email addresses for their own purpose, and some people are unwilling to display their email address and phone number on an SNS. Moreover, personal information crawling has gradually restricted based on laws in the aspect of the privacy protection. An analysis system of mobile devices has an advantage of analyzing without transfer personal private data externally.

As a result, a sentiment analysis of personal profiles presents a problem in that the characteristics are unstructured and there is a lack of data for fine-grained analysis from the data perspective. In addition, from a media perspective, application integration and integration reference are required to compensate for the problem of specialized single media and concentrated user groups. Finally, an analytical method that would ensure the protection of privacy in the case of a server-side analysis is required.

3 Related works

A sentiment analysis is used to analyze the subjective emotions of people regarding products [Ilarri, Hermoso, Trillo-Lado et al. (2015)]. A number of research efforts on social media have been carried out to analyze user interests because the user's current feelings are expressed well through SNS [Kim, Bak and Oh (2012); Ha, Back and Ahn (2015); Etter, Colleoni, Illia et al. (2018); Zimbira, Abbasi, Zeng et al. (2018)]. From the perspective of the granularity degree of the target, although previous analyses were aimed at the document and sentence level, sentiment analyses are being increasingly fine-grained to the feature level. In addition, structural information for each feature is required for building a hierarchical structure. These hierarchical approaches have the advantage that the upper layer can be considered as an alternative in the absence of an equivalent layer. Hierarchical approaches for fine-grained feature-level data can enhance the understanding in terms of various and specific features. Riloff et al. [Riloff, Patwardhan and Wiebe (2006); Yu, Zha, Wang et al. (2011)] have proposed to construct and analyze the knowledge structure of the features which is extracted from data. Cambria et al. [Cambria, Poria, Hazarika et al. (2018)] propose SenticNet 5 which discover conceptual primitives using context embeddings. Majumder et al. [Majumder, Hazarika, Gelbukh et al. (2018)] utilize context modeling and commonsense knowledge embedding in sentiment analysis. To reduce the difficulty of a term owing to a fine granularity of feature, ontology representation scheme can be used to construct hierarchical relation of fine-grained features.

A sentiment ontology was proposed for sentiment learning of product reviews, and it represents a sentiment value using features with a tree structure [Wei and Gulla (2010)]. Recently, researches on sentiment ontology are extended to multi-language [Liu, Jou, Chen et al. (2016)] and cross-media [Cao, Ji, Lin et al. (2016)], and widely used to analyze social sentiment on government intelligence domain [Kumar and Joshi (2017)]. Dragoni et al. [Dragoni, Poria and Cambria (2018)] propose OntoSenticNet which consists of commonsense ontology in sentiment analysis.

In the aspect of sentiment calculation, sentiment analysis can be used for analyzing the polarity of the target according to the intensity, which is divided into positive, neutral, or negative polarity. SentiWordNet contains the emotional information of each word and can be used to describe the sentiment polarity and quantify the intensity of polarity between 0 and 1 [Esuli and Sebastiani (2007)]. A negative opinion describes as a negative number and a positive opinion describes as a positive number. In addition, a zero represents a neutral opinion such as specification and facts. SentiWordNet was updated to version 3.0, and further studies applied a correlation calculation to improve the performance [Baccianella, Esuli and Sebastiani (2010); Kim, Bak and Oh (2012); Hung and Lin (2013)]. Recently, SentiWordNet is utilized in sensitive embedding [Dorkar and Joshi (2017)] and semi-supervised approaches, which incorporate lexicon-based methodology with machine learning [Khan, Qumar and Bashir (2017)].

A standard parameter is a need to integrate the medium as a reference for a plurality of different media and the temporal information can be used for the reference parameter. Temporal information indicating the ontology has been studied, such as temporal RDF (Resource Description Framework), OWL (Web Ontology Language) time, and temporal description logic. Temporal description logic is an extension of the description logic, which is

used to process the time representation with extra effort [Liu, Jiang, Sha et al. (2012)]. OWL time describes an interval form with a temporal entity class instance and displays the time of web page content and web services [Hung and Lin (2013)]. The methods for representing time mostly use the start and end times. Alternatively, time can also be represented using the start time and an interval. Ontological inferences have been conducted regarding temporal information [Baccianella, Esuli and Sebastiani (2010)]. In the aspect of the temporal influence, some studies assigned more weight to the latest contents. Previous research used a fixed constant to reduction ratio because it is not easy to define the influence of temporal weights. In the aspect of psychology perspective, the forgetting curve of Ebbinghaus can represent the effects of human memory according to time [Artale, Knotchakov, Wolter et al. (2013)].

Emotional information has a direct impact on personal privacy, and a large amount of emotional data is generated in a mobile device. The research which directly connects to the server as well as extracts data from the mobile was performed. These researchers analyzed the stored text, conversations, and user context information transmitted to the server [Milea, Frasinca and Kaymak (2008)]. A natural language analysis was conducted by connecting to a semantic assistant server that uses a natural language process (NLP) [Sateli, Cook and Witte (2013)] and a convolutional neural network (CNN) [Xu, Zhang, Xin et al. (2019)]. A natural language analysis and data mining techniques, which are mainly carried out on a PC, have also been carried out to enable analysis of mobile devices. There are some approaches to analyze data on the mobile devices such as AnalyticDroid, Weka-for-android, and TensorFlow Lite. Recently, Wang et al. [Wang, Zhang, Bao et al. (2018)] propose ARDEN which privacy preserving hybrid solution and Luong et al. [Luong, Xiong, Wang et al. (2018)] use multi-layer neural network for optimal auction on mobile blockchain environment. Currently, data mining analyses on mobile devices are mainly conducted on a server or processed using relatively small size data. However, it is possible to handle a large amount of data in accordance with the performance improvements of mobile CPUs and GPUs [Schrag (2012)].

The user profile is a method to analyze an individual's propensity to express a user's preference. Conventional user profile focuses on the user's external information such as height and weight. However, online user profiles have been studied focusing on the user preferences rather than on external information. In terms of data acquisition, there are methods that enable users to directly input their preferences such as 'like' on Facebook. Those explicit user profiles are, in the aspect of data acquisition, targeted and clear solution. However, it needs user efforts to fill out each question and cannot be reused to other applications. A feature-level fine-grained analysis would require a large amount of data to be entered for each domain. There is a certain degree of risk that the actual selection differs from the real favorite. Text mining method can be used to create a user profile with existing user data instead of direct input. Zheng et al. [Zheng, Li and Sangaiah (2018)] use neural word embedding for group user profile in social network and Degha et al. [Degha, Laallam, Said et al. (2018)] utilize human profile ontology in smart building domain.

The existing user profile presents a method for analyzing the user preferences by using the weight of the frequency based on the tags written by the user. Frequency-based approaches are one of the frequently used methods to create a user profile. For this purpose, studies Liu et al. [Liu, Chen, Tang et al. (2012); Griol, Molina and Callejas (2013); Park, Tickoo,

Narayanan et al. (2015)] were conducted using the term frequency and inverse document frequency (TF-IDF) and best matching 25 (BM25) to the folksonomy-based analysis. Alternatively, studies Nembhard et al. [Nembhard and Uzumeri (2000); Mezghani, Zayani, Amous et al. (2012)] were carried out for creating a user profile by using the friends of a friend (FOAF) based on the relationship of the social network. A structural study of user profile form is performed to create a user profile using the ontology. It has the advantage of creating a profile with a hierarchical form. Research to create a user profile on a social network through reasoning has also been carried out [Wang, Zhang, Bao et al. (2018)]. The study of a user profile was carried out for recommendations using twitter. In addition, rule-based reasoning with ontology is used to make student's profile [Nafea, Maglaras, Siewe et al. (2016)]. Recently, Uddin et al. [Uddin, Rakib, Haque et al. (2018)] extract rules to match the users' needs with heterogeneous knowledge sources. User preference on e-learning domain is described with fuzzy ontology, which consist with fuzzy concepts and relations [Shao, Yang and Ma (2017)].

4 General-purposed user profile system based on sentiment analysis

The mobile device has become the medium with the most personal information. Maintaining the privacy of such an analysis of an individual requires a novel mobile system that would allow an internal analysis of the raw mobile data. The basic flow of the general-purposed user profile system proposed in this paper comprises the six steps as shown in Fig. 1.

The sentiment analysis collected as a function of time is facilitated by the uniformity of time as a common time unit. Due to generate by the same clock on the mobile device, it can rearrange the data more easily with time sequence. After rearranging the data, the clustering module clusters data based on the temporal similarity in Step 1. Unfortunately, user data, which is based on an individual user experience, is sparse compared with the external public data from the web, so it is hard to create a data structure. To compensate for this limitation, we acquire external public data to create the data structure in Step 2. By comparison between external public data and individual private data, personal characteristics can be extracted in opposition to public opinions.

In Step 3, when the opinions of the public and private individuals regarding a common topic are compared, consensus and opposing opinions can be distinguished. An individual opinion that is contrary to public opinion forms a better representation of the individual's characteristics than an agreeing opinion.

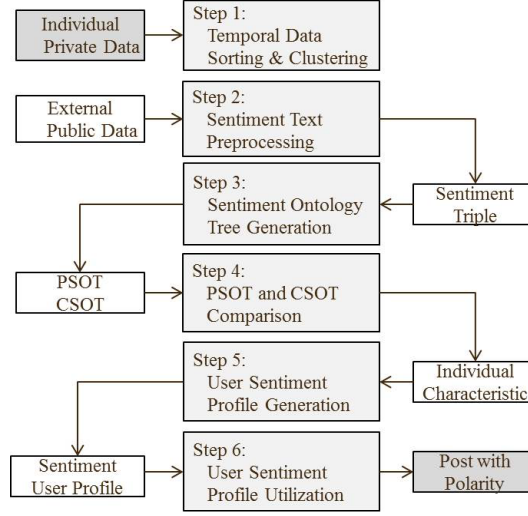


Figure 1: Proposed process of the general-purposed user profile system

In Step 4, the polarity of an instance constituting a collective sentiment ontology tree (CSOT) is calculated using the intensity and polarity of sentiment words that are defined in advance by SentiWordNet. The SentiWordNet contains the sentiment intensity and polarity of each sentiment word, and is also used for a feature-level analysis. As shown in Eq. (1), the polarity value of the feature is composed of a sum of the product of the intensity, weight, and number of sentiment words. The sentimental value of the i_{th} feature f_i is calculated by adding all of the n_{csw} weighted sentiment words of the i_{th} feature f_i .

$$S_{csot}(f_i) = \sum_{j=1}^{n_{csw}} N_j \times W_j \times I_j \quad (1)$$

where $S_{csot}(f_i)$ denotes the CSOT polarity value of the i_{th} feature f_i , n_{csw} is the number of sentiment words of the i_{th} feature f_i of the CSOT, N_j indicates the number of j_{th} sentiment word, and W_j is the weight for the j_{th} sentiment word. In addition, I_j means the intensity of the j_{th} sentiment word, which is extracted from SentiWordNet. When the calculated polarity value is negative about feature, it means data negative opinion about it. Conversely, positive value indicates a positive opinion, whereas a neutral opinion has a value of zero. The strength of the polarity is determined according to the degree of sentimental value.

For the consideration of temporal influence, it had been reduced weight with a constant decrease rate to assign a weight with respect to the latest content. However, in psychological terms, people's memories are reduced in proportion to the forgetting curve of Ebbinghaus [Nembhard and Uzumeri (2000)] instead of decreasing with a constant ratio. The formula of the forgetting curve by Ebbinghaus was used to calculate the weight according to the time, as shown in Eq. (2), where W_j indicates the temporal weight of the j_{th} sentiment words, t represents the difference between the target time and the current time, and parameters c and k are empirically determined values for the forgetting curve ($c=1.25$, $k=1.84$) used in [Nembhard and Uzumeri (2000)].

$$W_j = \frac{100k}{(\log t)^{c+k}} \quad (2)$$

The sentimental value of the CSOT regarding the i_{th} feature f_i is rewritten with a

combination of Eqs. (1) and (2), as shown in Eq. (3).

$$\begin{aligned}
 S_{csot}(f_i) &= \sum_{j=1}^{n_{csw}} N_j \times W_j \times I_j \\
 &= \sum_{j=1}^{n_{csw}} N_j \times \frac{100k}{(\log t)^{c+k}} \times I_j \\
 &= \sum_{j=1}^{n_{csw}} \frac{100k \times N_j \times I_j}{(\log t)^{c+k}}
 \end{aligned} \tag{3}$$

The sentimental value of the personal sentiment ontology tree (PSOT) regarding the i_{th} feature f_i can be calculated using Eq. (4) below, where n_{psw} indicates the number of sentiment words about the i_{th} feature f_i of a PSOT. The other parameters are the same as in Eq. (3).

$$S_{psot}(f_j) = \sum_{j=1}^{n_{psw}} \frac{100k \times N_j \times I_j}{(\log t)^{c+k}} \tag{4}$$

In the aspect of the polarity and intensity, the sentiment value between PSOT and CSOT are compared with two Eqs. (5) and (6). Eq. (5) is used to calculate the polarity $P_{csotpsot}$ of the same feature f by multiplying the sentimental values of $S_{csot}(f_i)$ in Eq. (3) with $S_{psot}(f_i)$ in Eq. (4).

$$P_{csotpsot}(f_i) = S_{csot}(f_i) \times S_{psot}(f_i) \tag{5}$$

When $P_{csotpsot}(f_i)$ has a negative value, f_i is defined as a major element because it indicates a contrast between public and individual preferences. When $P_{csotpsot}(f_i)$ has a positive value, it has been defined as a minor element, which has the same polarity. The value of deviation $D_{csotpsot}(f_i)$ between CSOT $S_{csot}(f_i)$ and PSOT $S_{psot}(f_i)$ can be calculated based on the absolute value of the sentiment difference regarding the same feature f_i in Eq. (6).

$$D_{csotpsot}(f_i) = |S_{csot}(f_i) - S_{psot}(f_i)| \tag{6}$$

The value of $D_{csotpsot}(f_i)$ indicates the difference in intensity. When the value of $D_{csotpsot}(f_i)$ is large, the difference in intensity is significant. The intensity values are used to rank the influence of each of the major and minor elements.

5 System architecture of general-purposed user profile system

As depicted in Fig. 2, the proposed general-purposed user profile system consists of four main modules: a data acquisition module, collective sentiment ontology tree (CSOT) module, personal sentiment ontology tree (PSOT) module, and a user-centric sentiment profile module. Along the horizontal axis, the left side of the diagram represents the handling of individual private data on mobile devices and the right side represents the mass of external public data on the web. The vertical axis is configured with temporal sequence.

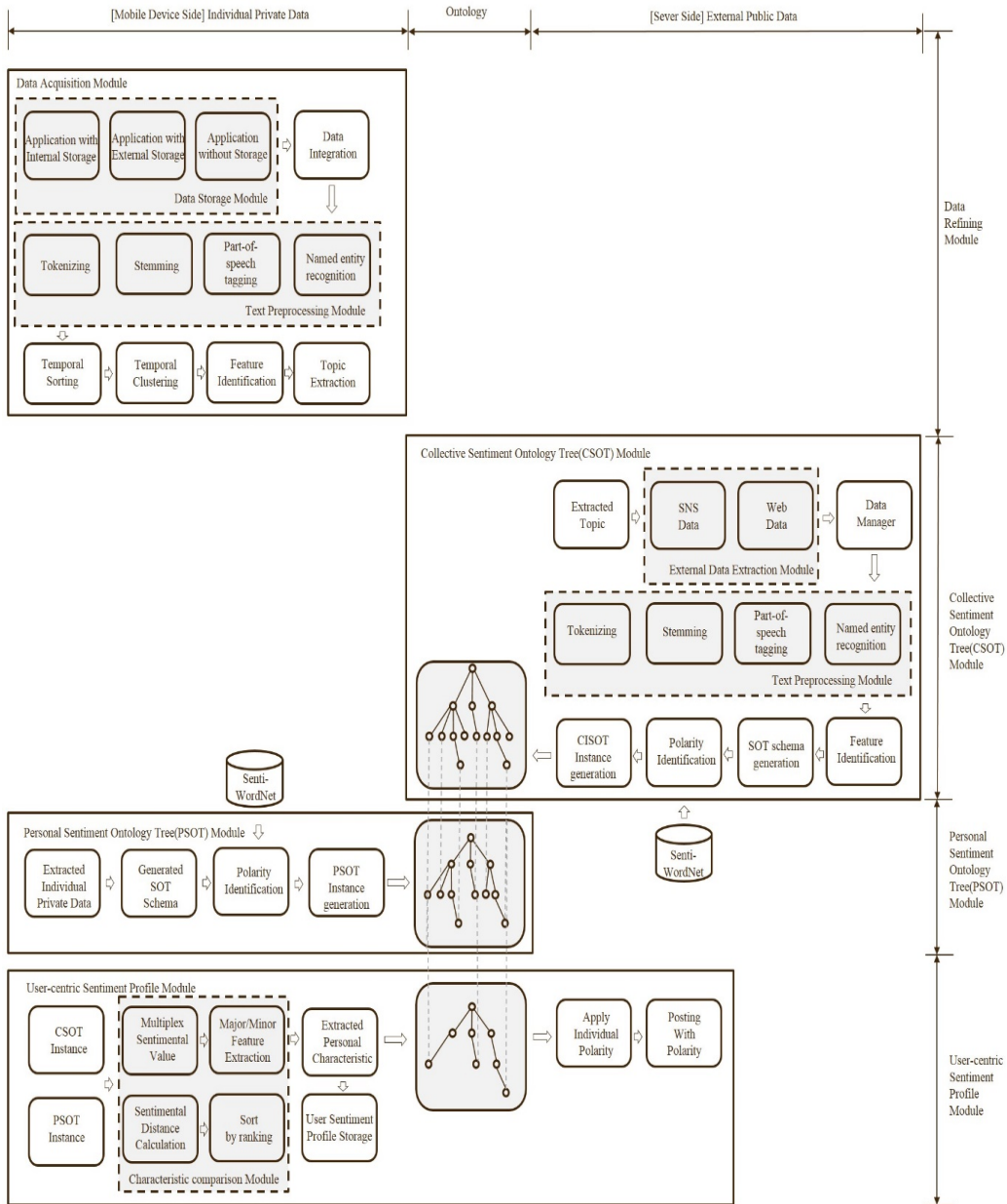


Figure 2: Architecture of general-purpose user profile system

Personal data, which is scattered in diverse forms in various application on mobile devices is collected by the data acquisition module based on each individual user and sorted the data in chronological order. And then, temporal clustering module clusters the individual private data with the time correlation. The topic is selected by using topic modeling on a grouped cluster and the conducted topic is used to generate a sentiment ontology tree schema with external public data. The proposed general-purpose user profile system uses SentiWordNet to extract each sentiment polarity and intensity with respect to the individual private data and external public data for a specific topic and features. Furthermore, the system generates PSOT instances of the individual private data and CSOT instances of the external public data by considering the temporal weight of the forgetting curve. By a comparison between CSOT and PSOT, the ranked major and minor element of

the individual characteristics are extracted as the user sentiment profile. A more detailed description about each module will be presented in the following subsections.

5.1 Data acquisition module

The data acquisition module is used to collect and integrate a variety of individual private data on mobile devices. After text preprocessing and temporal clustering, the module extracts the topics and features. The sub-modules of the data acquisition module consist of a data storage module, data integration module, text preprocessing module, temporal clustering module, and topic extraction module.

The data storage module is responsible for collecting data according to each of the properties of application. It is composed of an application with internal storage, an application with external storage, and an application with no storage based on the type of data acquisition of the application. For an application with internal storage, it is extracted directly from the underlying database as a bookmark and SMS message. It extracts data and time information from the application by sending intent or accessing application database. For an application with external storage, such as Facebook or Twitter, the system extracts the time and data from the response message to the server or application-provided API. For an application without storage, such as a telephone or video streaming service, it is difficult to analyze the text information without converting voice into text. There are some alternative analysis methods that transmit the data to the server and process the natural language.

The data integration module has a function of conversion into the same unit when the data and time units differ according to each application. The type of information processed by the data integration module is composed of both temporal and textual information. When the temporal information of an application is stored in an epoch time unit, the data integration module converts this into minutes during the temporal weight calculation process.

The text preprocessing module has the function of converting raw text to analyzable text data through four text preprocesses: tokenizing, part-of-speech (POS) tagging, stemming, and named entity recognition (NER), to analyze the text data extracted from each application.

The temporal clustering module sorts individual private data according to the time sequence in chronological order and generates clusters based on the relevance of the temporal correlation, as depicted in Fig. 3. The uniformity of such a time as a common time unit makes it easier to reorder the information generated by a mobile device and to integrate information of a similar time within a cluster. When the temporal difference is low, considering the similarity of the time, there are more similar themes and relevance.

The topic extracting module extracts the features for each cluster, playing the role of detecting a topic of interest to the users from the extracted features.

5.2 Collective sentiment ontology tree (CSOT) module

CSOT modules are used to extract the features and preferences of the public through the Web and SNS regarding a detected topic. It consists of an external data extraction module, a data integration module, a text preprocessing module, an SOT schema module, and a

CSOT instance module. The data integration and text preprocessing modules have the same functionality as those used in the data acquisition module.

The external data extraction module is used to extract external public data from the Web and an SNS. It is difficult to generate a data structure using mobile device data for a specific topic. The external data acquisition module gathers the data collected through Web and social network data searches on a particular topic as reference data. A point of view of data characteristics, Web pages, or blog data has an advantage of dealing with detailed information about the target, but it lacks sentimental and real-time information compared to the data on an SNS. Because SNS deals with a large amount of recent information, the data on such services have an advantage in terms of real-time and sentiment information. On the other hand, there is a limitation to generating complete structural data independently because the data size of an SNS is relatively small and scattered. This module extracts data from the Web and SNS for collecting various data and building the data structure.

A sentiment ontology tree schema module constitutes an SOT schema based on the extracted features after preprocessing the data from the Web and SNS. Each leaf node consists of a hierarchical layer of the extracted feature. This schema is used to create CSOT and PSOT instances in the next step.

A CSOT instance module generates a CSOT schema by adding sentiment words mentioned for the corresponding feature to the child node of the SOT schema. Based on the SOT schema obtained with respect to the topic, a CSOT instance indicates a sentiment instance of the public for each feature.

5.3 Personal sentiment ontology tree (PSOT) module

Whereas a CSOT instance indicates a sentiment instance of the external public data for each feature, a PSOT instance indicates a sentiment instance of individual private data for each feature. The sentiment words mentioned for the corresponding feature are added to the child node in the SOT schema.

The PSOT instance module is used to generate a PSOT instance using both individual private data from the data acquisition module and the schema generated by the SOT schema module. The method for calculating PSOT sentimental values is the same as for the CSOT. It can be obtained by multiplying the number, weight, and extracted intensity in SentiWordNet using Eq. (4). The PSOT instance shows the SentiWordNet sequence number and sentiment value of the sentiment words on visualization.

5.4 User-centric sentiment profile module

The user-centric sentiment profile module is used to extract the individual characteristics of the same features by comparing the value of the PSOT instance with the value of the CSOT instance. Two types of evaluation parameters are the polarity and deviation which are calculated using Eqs. (5) and (6) in Section 4. Applying the individual polarity module is concerned with the usage of sentiment user profile. When the topic of concern and the feature is shown when posting text, personal preferences of the topic can be added on posting a message or recommending goods to improve the understanding of a similar domain. For example, when someone posts a message *'this movie is quite boring'*

the system adds a footer such as ‘*prefer: action, adventure*’. When people click the “like” button for a camera lens on Facebook, the system could add the comment ‘*prefer: aperture weight*’ for the other users who read the review. The proposed system has the advantage in that new data can be analyzed based on the accumulated data generated on mobile devices. Finally, a user sentiment profile, which is based on the analyzed data for each topic, is used to construct an overall sentiment profile for an individual.

The characteristic comparison module has the function of comparing the value of the PSOT instance with the value of the CSOT instance. By the multiplication of the sentimental value of PSOT and CSOT instance, it separates major and minor features based on the difference in polarity using Eq. (5). For the difference in intensity, the module calculates the sentimental distance using Eq. (6), after which it uses ranking to sort the top-k features based on sentimental distance.

The user sentiment profile storage module constructs the sentiment user profile consisting of ranked major and minor features using the extracted polarity and intensity difference. The user preference is stored internally as the sentiment user profile for each topic. The sentiment user profile is accumulated with the extracted user preference and incremental analysis is available. The similarity calculation between topics stored in the user profile and new topics for analysis uses the stored user preference for estimating and preventing data sparsity for the new topic.

6 Experimental results

The experimental results of general-purposed user profile system is described in this section. First, the layered feature relative to the extracted topic are visualized as the results of CSOT schema. Furthermore, by using each feature based on the CSOT schema, the PSOT instance is illustrated sentiment expression in the form of a sentiment tree. And then the ranked major and minor features based on the polarity and the intensity between CSOT and PSOT are shown. Finally, we analyzed the results when the system applied the temporal weight of the forgetting curve. Figs. 3, 5 and 6 are visualized with AChartEngine and Fig. 4 is visualized by Radial Reingold Tilford Tree [Dorkar and Joshi (2017)].

In this study, mobile devices data were collected with 56,135 pieces for the feasibility of the system, including 31,336 Web histories with 24,799 cases of user information, which included 3,078 bookmarks, 16,483 KakaoTalk messages, 105 Facebook messages, 193 tweets, 30 Facebook posts, 566 emails, and 4,344 SMS/MMS messages.

Fig. 3 shows a representation of temporal clustering using the expected maximization clustering according to the similarity of the data collected from a mobile device during the period of May 20 to May 26 (see Step 1 on Section 4 for details). After pre-processing, temporal clustering using k-means clustering was conducted based on the similarity of the data collected from mobile devices. It shows five clusters with data on seven mobile device applications for a week-long period. The *x-axis* represents application name on mobile device and the *y-axis* represents the usage of application with epoch timestamp.

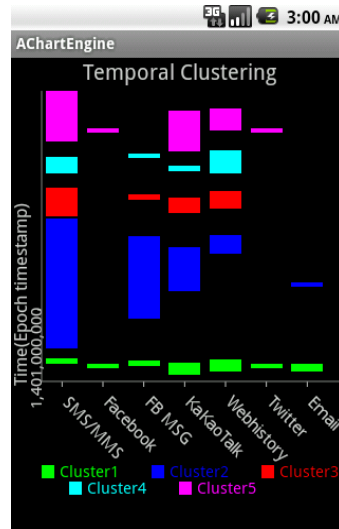


Figure 3: Temporal clustered application data

Fig. 4(a) shows the visualization of an SOT schema composed of 225 feature nodes and 10 upper-layer nodes (see Step 3 on Section 4 for details). As partially shown in Fig. 4(a), the schema of CSOT was constructed using 10,138 sentences of Web data for the extracted topic.

For example, price upper-layer node has seven lower-layer node such as sales, repair fee, and discount. As shown in Fig. 4(b), the instance of PSOT is composed of 2,847 sentiment words and 298 kinds of sentiment words as SentiWordNet. Feature ‘tops’ has four sentiment words such as comfortable, great, and good. Sentiment word ‘wide’ of feature ‘shirts’ describe with ‘wide#I@P_0.25_O_0.625_N_0.125’. ‘P_0.25’ means that sentiment word ‘wide’ has 0.25 positive value. ‘O_0.625’ represents 0.625 object values and ‘N_0.125’ means 0.125 negative values.

The difference between temporal weighted values and non-weighted values on period 4 was depicted in Tab. 1. G_{psot} indicates the PSOT value of non-weighted and G_{wpsot} means the temporal-weighted PSOT value. G_{csot} indicates the CSOT value for each features. In addition, the difference between the purchase history and user-centric sentiment profile analysis based on the point of purchase to describe the accuracy is shown in Fig. 5. It shows the influence of the proposed system with the Ebibinghaus temporal weight. The x -axis represents purchase periods and the y -axis represents the intensity between weight and non-weight of temporal information.

In experimental evaluation, the user-centric sentiment profile is constructed with utilizing accumulated user data for clothe domain. We calculated with the sentiment value of proposed methods which learned on each period for the purchased products. The proposed method of Ebibinghaus weights shows 24.9% better accuracy compared with the method which has not considered temporal weight over the 6th period.

Table 1: Difference between temporal weight and non-weighted on period 4

Feature	G_{psot}	G_{wpsot}	G_{csot}	$\frac{G_{psot}}{G_{csot}}$	$G_{wpsot}-G_{csot}$
Blue	0.044	0.045	0.161	-0.115	-0.116
Color	0.087	0.084	0.080	0.004	0.006
Fabric	0.116	0.112	0.085	0.027	0.031
Fit	0.156	0.151	0.143	0.008	0.013
Khaki	0.227	0.248	0.134	0.114	0.093
Length	0.045	0.044	0.107	-0.063	-0.062
Look	0.149	0.144	0.094	0.051	0.055
Material	0.061	0.059	0.107	-0.048	-0.047
Shirts	0.115	0.112	0.089	0.023	0.026
Blue	0.044	0.045	0.161	-0.115	-0.116
Color	0.087	0.084	0.080	0.004	0.006

Table 2: List of extracted major and minor features

Feature	$S_{psot}(F_i)$	$S_{csot}(f_i)$	$D_{csotpsot}(f_i)$	$P_{csotpsot}(f_i)$
Overall	-5.595	0.049	5.645	Major
Shape	-2.592	0.103	2.695	Major
Shoes	-1.183	0.129	1.312	Major
Front	-1.555	0.026	1.581	Major
Legopen	-0.174	0.052	0.226	Major
Size	0.366	-3.185	3.551	Major
Fabric	-0.532	10.201	10.733	Major
Color	-1.728	10.976	12.704	Major
Quality	1.037	15.962	14.926	Minor
Feel	0.691	7.878	7.187	Minor
Material	7.084	1.059	6.025	Minor
Ordinary	4.417	0.620	3.527	Minor
Price	3.110	0.465	2.645	Minor

Tab. 2 shows the list of extracted major and minor features with sentiment values which is calculated with equations on Section 4. Fig. 6 shows the result of the analysis of the sentiment difference with an instance of PSOT and CSOT.

The x -axis shows the feature list of the major and minor elements. The y -axis represents the intensity of the PSOT and CSOT instances for major and minor elements. The

sentimental value of the CSOT instances and PSOT instances are calculated using Eqs. (3) and (4), respectively, for same 86 common features. The value of the opposite polarity was selected to define a major feature, whereas the same polarity value was selected as a minor feature using Eq. (5).

After calculating the deviation using Eq. (6), the major and minor top-k can be determined. According to the experiment, the major features include the color, fabric, and size, and the minor features include the quality, feel, and material. These features are considered as sentiment user profile, which can represent individual preference on the clothing domain.

Tab. 3 describes top five features with negative and positive values of PSOT, CSOT and the proposed method. For example, overall and shape are negative features of PSOT. Color and fabric are major features of proposed method. Tabs. 4 and 5 show the accuracy of top five features of PSOT, CSOT and proposed methods during seven purchase period. The accuracy of Tab. 4 is calculated with major features. However, the accuracy of Tab. 5 considers major and minor features. Proposed method, which consider major features, shows better accuracy with 52.8% and 42.8% for PSOT and CSOT, respectively.

When minor features are also considered with major feature, Tab. 5 shows 41.0% and 18.6% improved for PSOT and CSOT, respectively. The reason of improvement reduction is that PSOT and CSOT considered minor features which are considered major features on proposed method.

Table 3: Top five features with negative and positive values

		PSOT		CSOT			
		Negative	Positive	Negative		Positive	
overall	-5.595	material	7.084	size	-3.185	quality	15.962
shape	-2.592	ordinary	4.417	waist	-1.524	color	10.976
color	-1.728	suit	3.501	calf	-0.465	fabric	10.201
front	-1.555	price	3.110	cheap	-0.439	feel	7.878
hoes	-1.183	quality	1.037	width	-0.297	material	1.059

Proposed Method			
Major		Minor	
color	12.704	quality	14.926
fabric	10.733	feel	7.187
overall	5.645	material	6.025
size	3.551	ordinary	3.527
shape	2.695	price	2.645

Table 4: Accuracy based on major features

	P1	P2	P3	P4	P5	P6	P7	Avg
PSOT	0.00	0.05	0.13	0.18	0.11	0.06	0.00	0.08
CSOT	0.14	0.05	0.00	0.18	0.00	0.06	0.20	0.09
Proposed	0.14	0.03	0.08	0.27	0.11	0.08	0.40	0.16

Table 5: Accuracy based on major and minor feature

	P1	P2	P3	P4	P5	P6	P7	Avg
PSOT	0.14	0.18	0.38	0.64	0.33	0.15	0.00	0.26
CSOT	0.29	0.23	0.21	0.64	0.33	0.21	0.60	0.36
Proposed	0.43	0.15	0.33	0.82	0.33	0.21	0.80	0.44

7 Conclusion

This paper proposes our general-purposed user profile system construct the user profile from individual private data for the user preference. To overcome data-sparsity of user data, it utilizes the available external public data for the feature of the sentiment ontology tree. From a data perspective, the purpose of compensating for the non-structured characteristics and the data shortage resulting from the fine-grained analysis is to utilize a CSOT schema and to compare a CSOT instance based on external public data of the same topic. For a media integration of various and biased application, user-centric integration was performed by temporal clustering as the reference of temporal relevance. For protecting the privacy when a server-based method is used, methods using the data of the collective ontology tree were proposed that would not require transmitting personal data. The sentiment user profile is constructed with ranked major and minor features for the specific domain by comparing the instance of the CSOT with the instance of the PSOT using the temporal weight based on the forgetting curve of Ebbinghaus. The sentiment user profile is incrementally accumulated for diverse domains in the mobile device. In the future research, we plan to conduct a study on the characteristic of the group of people. We also plan to conduct research on ways to improve the performance of the system through parallelism.

Acknowledgement: This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2019-0-00231, Development of artificial intelligence based video security technology and systems for public infrastructure safety).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- Artale, A.; Kontchakov, R.; Wolter, F.; Zakharyashev, M.** (2013): Temporal description logic for ontology-based data access. *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 711-717.
- Baccianella, S.; Esuli, A.; Sebastiani, F.** (2010): SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the Seventh conference on International Language Resources and Evaluation*, pp. 2200-2204.
- Bird, C.; Gourley, A.; Devanbu, P.; Gertz, M.; Swaminathan, A.** (2006): Mining email social networks. *Proceedings of the 2006 International Workshop on Mining Software Repositories*, pp. 137-143.
- Brusilovsky, P.; Stock, O.; Strapparava, C.** (2000): *Adaptive Hypermedia and Adaptive Web-based Systems*. Springer.
- Cai, Y.; Li, Q.** (2010): Personalized search by tag-based user profile and resource profile in collaborative tagging systems. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pp. 969-978.
- Cambria, E.; Poria, S.; Hazarika, D.; Kwok, K.** (2018): SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. *Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 1795-1802.
- Cao, D.; Ji, R.; Lin, D.; Li, S.** (2016): A cross-media public sentiment analysis system for microblog. *Multimedia Systems*, vol. 22, no. 4, pp. 479-486.
- Degha, H. E.; Laallam, F. Z.; Said, B.; Saba, D.** (2018): Onto-SB: human profile ontology for energy efficiency in smart building. *Proceedings of 2018 3rd International Conference on Pattern Analysis and Intelligent Systems*, pp. 1-8.
- Dorkar, R. A.; Joshi, S. S.** (2017): Sentiment sensitive embeddings with SentiWordNet lexical database based cross-domain sentiment classification. *International Journal of Advance Research in Computer Science and Management Studies*, vol. 5, no. 8, pp. 86-93.
- Dragoni, M.; Poria, S.; Cambria, E.** (2018): OntoSenticNet: a commonsense ontology for sentiment analysis. *IEEE Intelligent Systems*, vol. 33, no. 3, pp. 77-85.
- Esuli, A.; Sebastiani, F.** (2007): SentiWordNet: a high-coverage lexical resource for opinion mining. *Evaluation*, pp. 1-26.
- Etter, M.; Colleoni, E.; Illia, L.; Meggiorin, K.; D'Eugenio, A.** (2018): Measuring organizational legitimacy in social media: assessing citizens' judgments with sentiment analysis. *Business & Society*, vol. 57, no. 1, pp. 60-97.
- Griol, D.; Molina, J. M.; Callejas, Z.** (2013): Providing personalized Internet services by means of context-aware spoken dialogue systems. *Journal of Ambient Intelligence and Smart Environments*, vol. 5, no. 1, pp. 23-45.
- Ha, I.; Back, B.; Ahn, B.** (2015): Mapreduce functions to analyze sentiment information from social big data. *International Journal of Distributed Sensor Networks*, vol. 2015, pp. 1-11.
- Hung, C.; Lin, H. K.** (2013): Using objective words in SentiWordNet to improve sentiment classification. *IEEE Intelligent Systems*, vol. 28, pp. 47-54.

- Ilarri, S.; Hermoso, R.; Trillo-Lado, R.; Rodríguez-Hernández, M. D. C.** (2015): A review of the role of sensors in mobile context-aware recommendation systems. *International Journal of Distributed Sensor Networks*, vol. 2015, pp 1-30.
- Khan, F. H.; Qamar, U.; Bashir, S.** (2017): A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. *Knowledge and Information Systems*, vol. 51, no. 3, pp. 851-872.
- Kim, S.; Bak, J.; Oh, A. H.** (2012): Do you feel what I feel? Social aspects of emotions in Twitter conversations. *Proceedings of the Sixth International AAI Conference on Weblogs and Social Media*, pp. 495-498.
- Kumar, A.; Joshi, A.** (2017). Ontology driven sentiment analysis on social web for government intelligence. *Proceedings of the Special Collection on eGovernment Innovations in India*, pp. 134-139.
- Liu, B.** (2010): Sentiment analysis and subjectivity. *Handbook of Natural Language Processing*. Taylor and Francis Group.
- Liu, B.; Jiang, Y.; Sha, F.; Govindan, R.** (2012): Cloud-enabled privacy-preserving collaborative learning for mobile sensing. *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pp. 57-70.
- Liu, H.; Jou, B.; Chen, T.; Topkara, M.; Pappas, N. et al.** (2016): Complura: exploring and leveraging a large-scale multilingual visual sentiment ontology. *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, pp. 417-420.
- Liu, P.; Chen, Y.; Tang, W.; Yue, Q.** (2012): Mobile weka as data mining tool on android. *Advances in Electrical Engineering and Automation*, pp. 75-80.
- Luong, N. C.; Xiong, Z.; Wang, P.; Niyato, D.** (2018): Optimal auction for edge computing resource management in mobile blockchain networks: a deep learning approach. *Proceedings of 2018 IEEE International Conference on Communications*, pp. 1-6.
- Majumder, N.; Hazarika, D.; Gelbukh, A.; Cambria, E.; Poria, S.** (2018): Multimodal sentiment analysis using hierarchical fusion with context modeling. *Knowledge-Based Systems*, vol. 161, pp. 124-133.
- Mezghani, M.; Zayani, C. A.; Amous, I.; Gargouri, F.** (2012): A user profile modelling using social annotations: a survey. *Proceedings of the 21st International Conference on World Wide Web*, pp. 969-976.
- Milea, V.; Frasincar, F.; Kaymak, U.** (2008): Knowledge engineering in a temporal semantic web context. *Proceedings of the Eighth International Conference on Web Engineering*, pp. 65-74.
- Nafea, S.; Maglaras, L.; Siewe, F.; Smith, R.; Janicke, H.** (2016): Personalized students' profile based on ontology and rule-based reasoning. *EAI Endorsed Transactions on e-Learning*, vol. 3, no. 12, pp. 1-18.
- Nembhard, D. A.; Uzumeri, M. V.** (2000): Experiential learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics*, vol. 25, no. 4, pp. 315-326.

- Park, M. S.; Tickoo, O.; Narayanan, V.; Irwin, M. J.; Iyer, R.** (2015): Platform-aware dynamic configuration support for efficient text processing on heterogeneous system. *Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition*, pp. 1503-1508.
- Park, S. M.; Baik, D. K.; Kim, Y. G.** (2016): Sentiment user profile analysis based on forgetting curve in mobile environments. *Proceedings of the 15th IEEE International Conference on Cognitive Informatics & Cognitive Computing*, pp. 207-211.
- Park, S. M.; Kim, Y. G.; Baik, D. K.** (2016): (Poster) Sentiment user profile system based on polarity comparison. *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services Companion*, pp. 142.
- Riloff, E.; Patwardhan, S.; Wiebe, J.** (2006): Feature subsumption for opinion analysis. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 440-448.
- Sateli, B.; Cook, G.; Witte, R.** (2013): Smarter mobile apps through integrated natural language processing services. *Proceedings of International Conference on Mobile Web and Information Systems*, pp. 187-202.
- Schrag, R.** (2012): Best-practice time point ontology for event calculus-based temporal reasoning. *Proceedings of CEUR Workshop*, pp. 28-34.
- Shao, J.; Yang, X.; Ma, C.** (2017): Research on fuzzy recommendation system based on user profile. *Open Journal of Social Sciences*, vol. 5, no. 10, pp. 1-11.
- Stoyanov, V.; Cardie, C.** (2008): Topic identification for fine-grained opinion analysis. *Proceedings of the 22nd International Conference on Computational Linguistics*, pp. 817-824.
- Sugiyama, K.; Hatano, K.; Yoshikawa, M.** (2004): Adaptive web search based on user profile constructed without any effort from users. *Proceedings of the 13th International Conference on World Wide Web*, pp. 675-684.
- Uddin, I.; Rakib, A.; Haque, H. M. U.; Vinh, P. C.** (2018): Modeling and reasoning about preference-based context-aware agents over heterogeneous knowledge sources. *Mobile Networks and Applications*, vol. 23, no. 1, pp. 13-26.
- Vallet, D.; Cantador, I.; Jose, J.** (2010): Personalizing web search with folksonomy-based user and document profiles. *Advances in Information Retrieval*, pp. 420-431.
- Wang, J.; Zhang, J.; Bao, W.; Zhu, X.; Cao, B. et al.** (2018): Not just privacy: improving performance of private deep learning in mobile cloud. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2407-2416.
- Wei, W.; Gulla, J. A.** (2010): Sentiment learning on product reviews via sentiment ontology tree. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 404-413.
- Whitelaw, C.; Garg, N.; Argamon, S.** (2005): Using appraisal groups for sentiment analysis. *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, pp. 625-631.

Xu, F.; Zhang, X.; Xin, Z.; Yang, A. (2019): Investigation on the Chinese text sentiment analysis based on convolutional Neural Networks in deep learning. *Computers, Materials & Continua*, vol. 58, no. 3, pp. 697-709.

Yu, J.; Zha, Z. J.; Wang, M.; Wang, K.; Chua, T. S. (2011): Domain-assisted product aspect hierarchy generation: towards hierarchical organization of unstructured consumer reviews. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 140-150.

Zheng, J.; Li, D.; Sangaiyah, A. K. (2018): Group user profile modeling based on neural word embeddings in social networks. *Symmetry*, vol. 2018, no. 10, pp. 1-23.

Zimbra, D.; Abbasi, A.; Zeng, D.; Chen, H. (2018): The state-of-the-art in Twitter sentiment analysis: a review and benchmark evaluation. *ACM Transactions on Management Information Systems*, vol. 9, no. 2.