Micro Auto Blogging by Using Granular Tree Based Context Model and AHP

Il-Kyoung Kwon and Sang-Yong Lee*

Division of Computer Science and Engineering, Kongju National University , 1223-24 CheonAn-Daero CheonAn-Si ChungCheongNam-Do, Korea
psent95@kongju.ac.kr, sylee@kongju.ac.kr

Abstract

It is necessary to have blogging technology which automatically figures out context information from the user in order to provide efficient micro blogging service for users in the social network service (SNS) environment. This paper proposes a context-based micro blogging service which considers the situations and emotions of users. To achieve this, a preprocessing method has been modeled based on users' location and time context by using a granular tree, and Naive Bayes Classification has been adopted to assess a user's behavior on the basis of a modeling context. In addition, a questionnaire has been administered to gain information about a user's emotion in different situations and which considers location by using the Analytic Hierarchy Process (AHP). Based on this, a blog-able single sentence generation and auto blogging have been performed. The evaluation result of blogging a sentence and user's emotional information shows 85.4% and 82.6% of accuracy, respectively; therefore, the proposed context modeling method for auto blogging is both efficient and effective.

Keywords: *Granular Tree, Bayes Classification, Analytic Hierarchy Process, Micro blog*

1. Introduction

Micro blogging service is becoming popular in the mobile environment due to the effect of SNS that blogs simple appreciation, for one. Micro blogging service is a type of blog that people are capable of communicating with many people using one or two simple and short messages. It enables users to send and receive information with short text, and the information is featured with real time update. As such, users can share information conveniently with other users by blogging his/her behaviors or location in different periods through a micro blog.

This paper proposes a context-based auto blogging system which takes users' emotion into account and considers positions and time contexts. To achieve this, a preprocessing method has been modeled based on users' location and time context by using a granular tree. The Naive Bayes Classification has been adapted to infer a user's behavior on the basis of a modeling context. In addition, a questionnaire has been administered to implement a blogging service about a user's emotion in different situations and which has evaluated a user's emotion in different places though the Analytic Hierarchy Process (AHP). Based on a user's context information and a user's emotion in different contexts, a 4W1E (Who, Where, When, What, Emotion) structure has been used to generate blog-able simple texts and implement auto blogging.

* Corresponding Author

2. Related Researches

2.1. Granular Computing

Granular Computing is a theoretical method to effectively use information granules, such as class, cluster, subset, group, or section, in order to build an effective calculating model to solve complicated applied problems of massive data, information, and knowledge. A granular tree which has hierarchical structure is used as an expression of granulation. Kocaball, in a research related to context matching, processed granules of range context located in two dimensions, using a hierarchical granular tree of 6 levels as a method of location matching [1]. Makkonen, in addition, used a hierarchical granular tree of 5 levels in processing information to measure interregional similarity of geographical ontology [2].

2.2. Naïve Bayes Classification

Naive Bayes Classification, a stochastical model assuming independences between natures based on Bayes' rule, is well-known in demonstrating a remarkable inference among classifications in spite of its easy learning method [3].

Under the assumption that each event is independent, Bayes' rule may be expressed as Formula 1. F denotes various forms of user input context, and this is an expression of probability of certain behavior (C) of the user under the condition where the input context occurs[4].

$$p(C \mid F_1,...,F_n) = \frac{p(C)\,p(F_1,...,F_n \mid C)}{p(F_1,...,F_n)} \tag{1}$$

Assuming Hypothesis F of Formula 1 is conditionally independent, the results of Naive Bayes Classification may be attained by following the process of Formula 2 and multiplying every probability of Hypothesis F.

$$p(C, F_1,...,F_n) \propto p(C)\,p(F_1 \mid C)\,p(F_2 \mid C)\cdots \propto p(C)\prod_{i=1}^{n} p(F_i C) \tag{2}$$

This study has adopted the Naive Bayes Classification to predict users' schedule at a specific time and location.

2.3. AHP

AHP, which is Analytic Hierarchy Process, is one of the backing methods for decision-making that catches an assessor's knowledge, experience, and intuition on the basis of his/her preferences through one-to-one comparison between elements that form a hierarchical structure. This method is well-applied in the field of decision-making, thanks to its theoretical simplicity, clarity, and ease of application. It especially provides a way to decide a priority in measuring intangibles and to analyze sensitivity upon changes in information involving decision-making.

Figure 1 presents the AHP process. The 1st and 2nd stages in Figure 1 show that the objects of analysis are organized in hierarchical order and scale for each evaluation criterion.

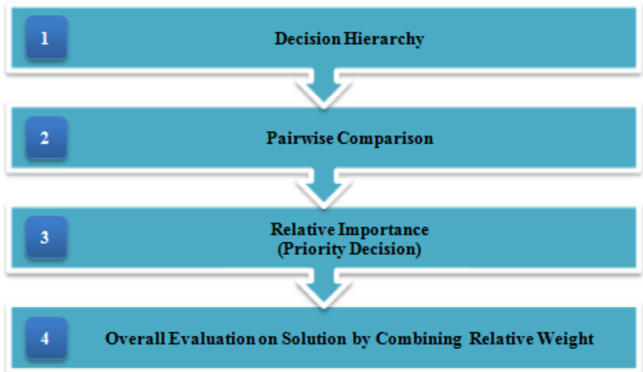


Figure 1. AHP Process

In the 3rd stage, a pair comparison is carried out to set the priority of each element. To do this, a judge matrix is prepared and overall priority is determined in changes of judges by comparing as many pairs as possible. Based on a determined priority, pairwise comparison A is presented as a diagonal matrix:

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & 1 & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & 1 & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & 1 & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 1 \end{bmatrix}$$

$$a_{ij} = \frac{1}{a_{ij}}, a_{ij} = 1, \forall i$$

The relative importance of matrix elements can be defined with Satty's 9-point scale. Here, points 1-9 indicate that one action is preferred to other actions by experience and decision; 1 means "similar," 3 means "slightly important," 5 is "important," and 9 is "highly important," whereas 2, 4, 6, and 8 are middle points of the abovementioned points. This study has adopted the AHP in order to infer relative importance between a specific context obtained from a questionnaire and languages expressed emotionally corresponding to the context, carrying out a sensitivity analysis on how changes of context can affect changes in emotion [5].

3. Micro Auto Blogging System

This chapter describes a micro auto blogging system using granular tree-based context mode and AHP. This system consists of 6 modules as seen in Figure 2: sensing module, user schedule reasoning module, user behavior reasoning module, emotion decision module, and sentence generation module. sensing module collects user input information including GPS coordinates, app in use, and schedule information and then sends them to a preprocessing module. Context preprocessing module is in charge of the preprocessing of location and time context of users, whereas user schedule reasoning module is in charge of reasoning user schedule on the basis of the Naive Bayes Classification by using location in different locations and time contexts of users. User behavior reasoning module is in charge of reasoning the detailed behaviors of a user in different assessed user schedule by using a context tree. Emotion decision module decides by using AHP emotional information what a user may feel in different contexts, whereas sentence generation module carries out simple sentence generation and real time auto blogging on the basis of 4W1E (Who, Where, When, What, Emotion).

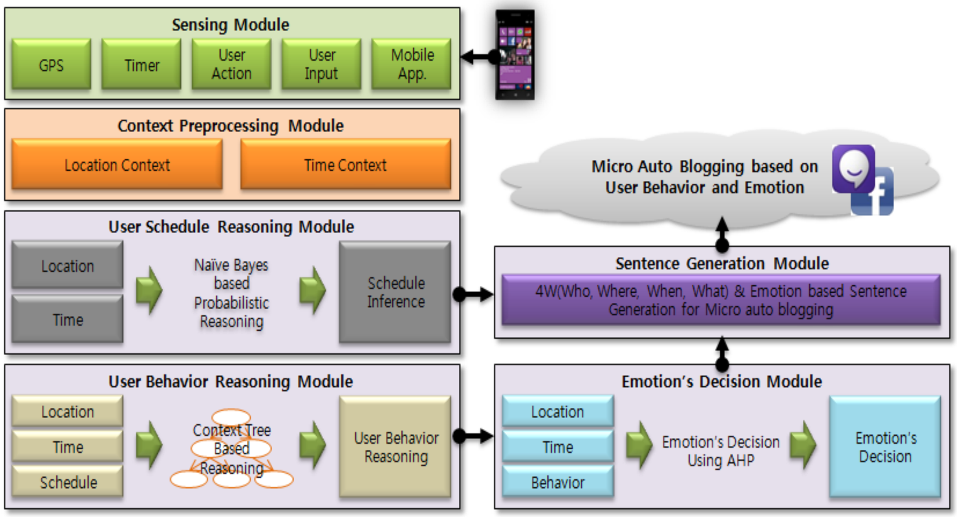


Figure 2. System Structure

3.1. Context Collecting

The types of contexts consist of received GPS coordinate, value of acceleration sensor, present date and time, *etc.* These contexts are input manually by users or automatically collected when regular time interval and movement exceeding set distances approach the designated place.

3.2. Preprocessing and Combining of Contexts

3.2.1. Preprocessing of Time Context: In this study, time context is preprocessed at a weekly interval, as Picture 2 shows, to make the search of time pattern convenient.

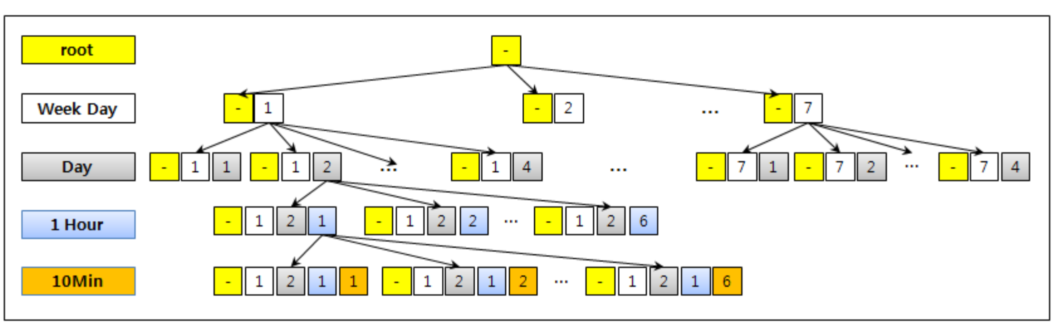


Figure 3. Preprocessing Time Context by Using Granular Tree

Time context is composed of 5 levels, including root that has the value of "-", to enable an index a week into 1,008 time units. When searched with "weekly," "daily," and "hourly" units in these time patterns, frequency and location of index number can be utilized by simply checking whether parts of patterns are matched.

3.2.2. Preprocessing of Location Context: Location context is presented as a hierarchical granular tree composed of 7 levels containing root as seen in Figure 4. The indication of users' major location and similarity scaling are performed by a hierarchical tree with digitization of real location context, which is similar to a granular tree.

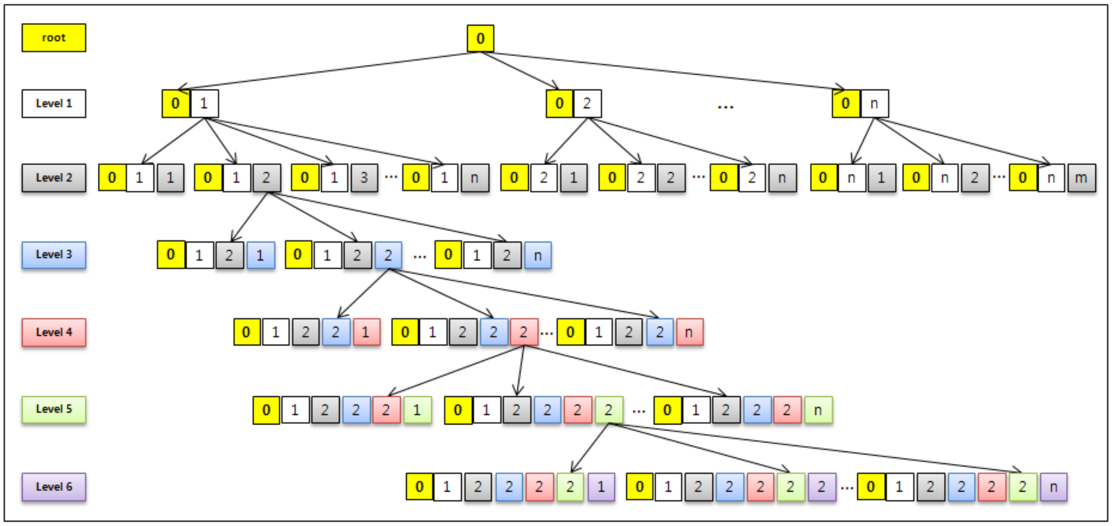


Figure 4. Preprocessing Location Context by Using Granular Tree

The tree in Figure 4 presents a location context with 6 similarity levels based on distance as seen in Table 1. Therefore, each level in the tree indicates similarity in different distances and the similarity is presented with a value of 0 and 1. However, similarity in Table 1 is set randomly in consideration of the experimental environment.

Table 1. Similarity of Granularity

Location	**Similarity level**	***Radius(meter)***	**Similarity**
District	1	2000m or more	0.1
Neighborhood	2	300~1999	0.2
Street	3	100~299	0.5
Building	4	50~99	0.8
Floor	5	30~50	0.95
Room	6	Within 10m	1.0

Table 2 presents examples of similarity scales between the queried location about location context and the target location. To measure this, first of all, queried location and target location are compared sequentially. When two locations are perfectly matched, calculate a value for matched section. Afterwards, compare the similarity table in Table 1 and find out the similarity level corresponding to the value and reflect it. When all the digits of a queried location context are matched with or contained in target contexts, it becomes "Y" in the satisfaction field and a similarity is set at a maximum value of 1 as seen in Table 2[1].

Table 2. Examples of Queries about Location Context

No	Queries	Target	Matched digit -1 (similarity level)	*Satisfaction*	Similarity
1	**022**31	**022**2162	2	N	0.2
2	**02231**	**02231**62	**4**	**Y**	**1.0**
3	**0222**322	**0222**222	3	N	0.5
4	**0222322**	**0222322**	**6**	**Y**	**1.0**

3.2.3. Combining Time and Location Context: As Figure 5 indicates, time context, location context, and schedule context are combined and expressed while a unit of combined context is expressed as 15 bytes long.

However, data storage for presenting each digit of location context cannot exceed 1 byte; it has constraints in practice as seen in Figure 5. Therefore, user schedule context is classified into 2 levels for storing. The 2 levels are for upper schedule and lower schedule, which correspond to user behavior. Each schedule is stored with frequency.

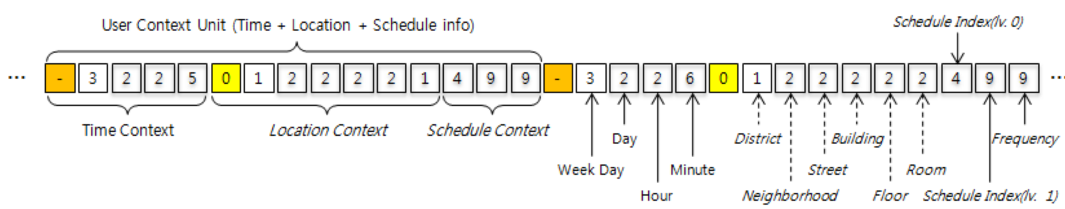


Figure 5. Combination of Time, Location, and Schedule Context

Structural context, as in Figure 5, enables scope matching (of time unit, range of location, location matching per time unit, *etc*.) simply by adjusting the digits of matching keywords.

4. User Schedule and Behavioral Reasoning

4.1. Schedule Reasoning

Schedule reasoning for the recognized context of user are performed by using the Naive Bayes Classification. User schedule is classified under 7 types with reference to GSS conducted by Statistics Canada. Each category is subclassified into the following: education such as assignment and seminar, social activity such as meeting or having meals, cultural activity such as going to a cinema, personal activity such as rest, dynamic hobby such as sports, and non-dynamic hobby such as book reading [6].

The Naive Bayes Classification will get a schedule pattern table composed of schedule information data, which is users' context, as learning data. A schedule pattern table consists of date and time data, weather, location, mobile app in use, and classification information. When newly performed schedule data is entered into a table, the probability of new data going into classification C can be calculated as seen in Formula 3. However, newly entered data is defined as I, overall number of schedule in the schedule table is defined as S, and classification for each schedule is defined as C.

When a schedule pattern table is defined as S, the probability P(C) of new data going into classification C can be calculated.

$$P(C_n) = \frac{P(S|C_n)}{P(S)} \times \frac{P(S|I.Time)}{P(S|C_n)} \times \frac{P(S|I.Location)}{P(S|C_n)} \times \ldots \frac{P(S|I.App)}{P(S|C_n)} \quad (3)$$

$$C = \{C_1 = assignment, C_2 = education, \cdots, C_7 = nondynamical\_hobby$$

4.2. User Behavior Reasoning

In practice, user behavior can be defined as users' detailed schedule; that is, when schedule for "sports activity" is actually "swimming," it can be said that a user is doing an activity called swimming (or a user is swimming). Therefore, user behavior reasoning means the process of reasoning about a detailed schedule out of a user's schedule.

Prior to user behavior reasoning, a context tree is organized by Classification C for a classified schedule. For this organization, a detailed pattern for a certain schedule is established as a context tree by using data in the existing schedule table. The example is presented in Figure 6. The context tree in Figure 6 includes a detailed schedule (Level 1) of education schedule (Level 0) and the location (Level 2), time (Level 3), and frequency of each behavior (F) that has been executed.

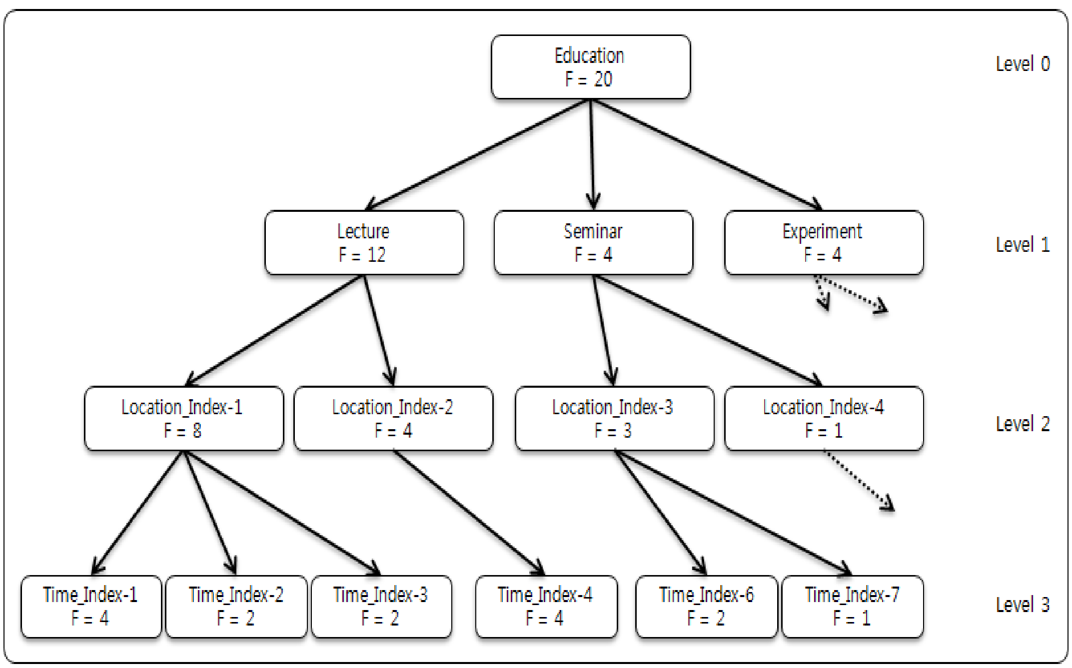


Figure 6. Examples of Context Trees in an Education Schedule

In order to select a detailed recommended schedule, similarity is measured on the basis of location and time context. First, it is assumed that a user's schedule context is organized as tree and frequency (F) for the node of each tree is determined. The points of each node in the context tree by using Formula 4 are obtained and the weighted value as many as the levels of the tree for the size of point is imposed. Therefore, time context in the detailed schedule becomes the most important factor in a schedule recommendation. The points of every node in the context tree are obtained by using Formula 5. The maximum value obtained from Formula 6 is selected as the most probable schedule context, which is obtained by using Formula 7[7][8].

$$Score(Node_k) = Level(Node_k) \times \sqrt{Frequency(Node_k)} \tag{4}$$

$$P(S,V) = Score(Node_s) + P(parent(S),V) \tag{5}$$

$$P(Contexts,V) = \sum_{context=Category_1\cdots}^{n} P(Context,V) \tag{6}$$

$$Behavior = \max(P(Contexts,V_1)\cdots P(Contexts,V_n))\ each\ of\quad V \tag{7}$$

S : Schedule, n : No. of category, Context : Category, V : User behavior

5. Emotional Reasoning

This study also aims to gain emotional information which is difficult to obtain in a mobile environment; therefore, emotional information in a specific context is gained through matching activity. The emotional information obtained is used to present emotional information in the context recognized by users. To do this, 38 emotional languages are selected out of 48 ones in 10 groups of Emotion Annotation and Representation Language (EARL) to be used as emotional information [9].

5.1. Emotional Expression Language and Context Information

Up until now, infra technology to directly gain, reason, and learn emotional context has not sufficiently developed among mobile platform-related intelligent technology. Despite this environment, gaining environment context and activity context are not difficult and this study has adopted Weather, User Action, Mobile App, and Location contexts. Figure 7 presents the results of the questionnaire provided to 100 university students to figure out the correlation between specific recognizable context information and emotionally expressive languages. One example of questions used in this survey is “emotions you may feel in the cinema,” which asks general emotions for a specific location, weather, user behavior, when using specific mobile app, *etc*. The vertical axis of Figure 7 presents a group of emotionally expressive languages based on the EARL classification, whereas the horizontal axis presents the frequency of emotionally expressive languages selected by respondents.

As seen from the result in Figure 7, positive emotionally expressive languages including “Delight,"”Excitement," "Happiness," "Joy," and "Pleasure" in the “positive and lively” group shows high frequency in terms of the location and weather-related context information. For the “quiet positive” group, “Relaxed" and "Relieved" are selected as the location of related emotionally expressive languages. In particular, highly positive emotionally expressive languages are selected in most of the recognized contexts.

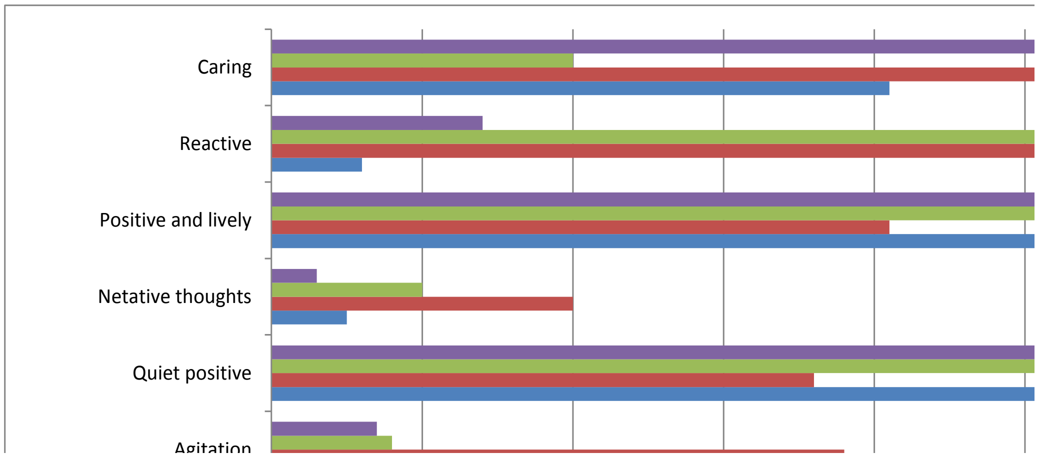
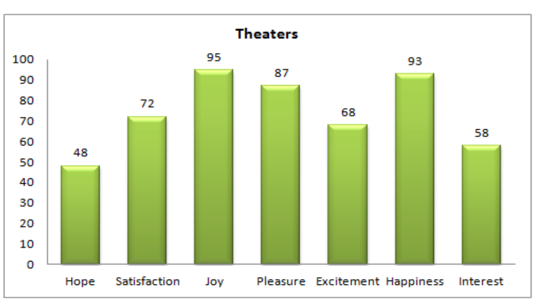


Figure 7. Selection Frequency of Positive Emotionally Expressive Languages from the Environmental and Behavioral Aspects

These positive or negative emotionally expressive languages can be used in the emotion-based personalized intelligent services. For example, in the “female returns home late at night" context, location, time, and transportation means in moving are negative (for example, “Fear”). It is possible to provide services such as executing music recommendation service or crime prevention-related application. Therefore, it is necessary to do a matching process between recognized context and emotionally expressive languages in order to provide personalized intelligent services using emotionally expressive languages.

5.2. Emotionally Expressive Languages and Context Matching

To match emotionally expressive languages with specific contexts, I have directly selected emotionally expressive languages that are felt in the context favored by the subject of the questionnaire (location, mobile app, user behavior, weather, *etc.*). Figure 8 presents emotionally expressive languages and their average frequencies in use are selected with higher rage in a specific favored location (cinema) and user action (web browsing). For example, the left chart in Figure 8 presents the frequency of emotionally expressive languages that are felt in a specific location called cinema by 100 respondents, which means an average of 91.6% (95, 87, 93 out of these 100 respondents) select “Joy", "Pleasure," and "Happiness." When the selection ratio of emotionally expressive languages in a specific location exceeds 70%, I have selected this as a universal emotional information on that specific context.



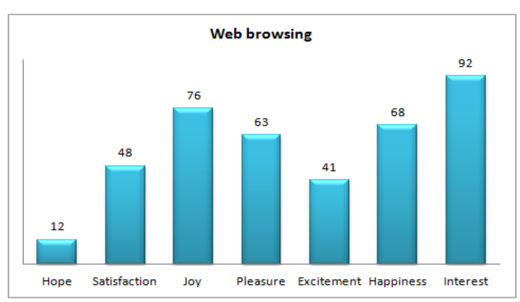


Figure 8. Emotional Expression Language Selected in Cinema (Left) and Web Browsing (Right)

From the survey result about emotionally expressive languages in a specific context, emotional information that is selectable in recognized context is obtained. Emotions favorably selected in a specific context are mostly matched with positive emotionally expressive languages. However, positive emotionally expressive languages felt in a specific location vary from person to other person and types of languages cannot be defined. Therefore, it is necessary to analyze to what extent a specific context affects the changes or degree of emotion. To do this, emotionally expressive languages in different contexts are analyzed for their sensitivity using AHP.

5.3. Analysis on the Relative Importance of Various Contexts Affecting Changes Of Emotion

Humans perform various and complex activities in different contexts, which means, they may exhibit the same behavior in the different places, or they may exhibit different behaviors in the same location in different times. Therefore, there is a need to analyze how different contexts and user behavior can affect the changes of emotions in people. To do this, I have established an AHP analysis model in consideration of various contexts and user behaviors and analyze 10 evaluators, who are not professional, but students selected randomly. To collect the opinions of 10 evaluators, I have adopted and analyzed the AHP results of each individual evaluator by using Formula 8.

There are 2 main methods of collecting the evaluation value of a group by using AHP; first, the opinions of evaluators are collected through discussion and voting and then a single pairwise comparison matrix is prepared based on the collected opinion. This is called "group evaluation method."

The second is called "value integration method," which collects each pairwise comparison matrix from a group member and then integrates the evaluation of the overall group and obtains the weight. This study has adopted a value integration method and integrated the element of pairwise comparison matrix by using the following formula: When the number of total evaluators are n and a_{ij} is the element in a pairwise comparison matrix that is assessed by kth evaluator, each element ($\overline{a_{ij}}$) of an integrated single pairwise comparison matrix can be obtained by the following formula 8.

$$\overline{a_{ij}} = \prod_{k=1}^{n} (a_{ijk})^{1/n} \tag{8}$$

Relative weight can be obtained for emotional changes in a complex context through an AHP analysis of contextual information. AHP result can be used as a practical index of emotion since it figures out how a specific context can affect the changes of emotion. Figure 9 presents the AHP hierarchical diagram to perform this.

Figure 10 presents a relative weight analysis result by using an expert choice [14] of context that may affect changes of emotion. The analysis result shows that emotion is sensitively changed in the order of schedule (L: .444), location (L: .321), and mobile app (L: .154) under execution. Furthermore, a detailed schedule shows that that emotion is sensitively changed in the order of sports (L: .231), shopping center (L: .167), and cinema (L: .105). In other words, if we can get users' current location, the schedule for the modeling of emotional expression to be used in micro auto blogging and universal emotion corresponding to a specific contextual information can be gauged with higher probability.

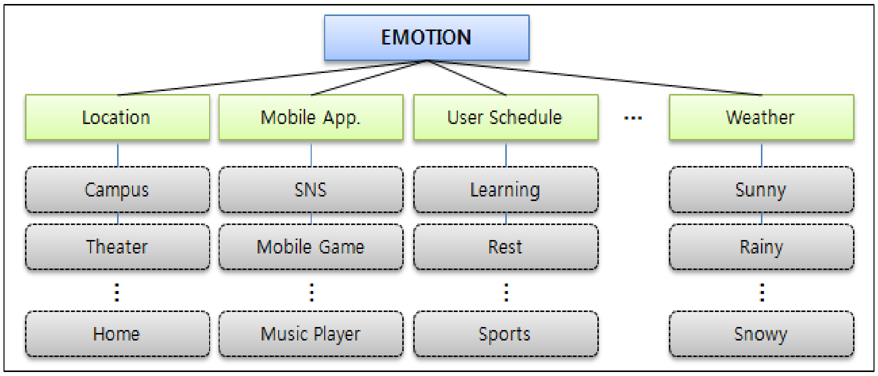


Figure 9. AHP for Analyzing Contextual Factors Affecting Emotions

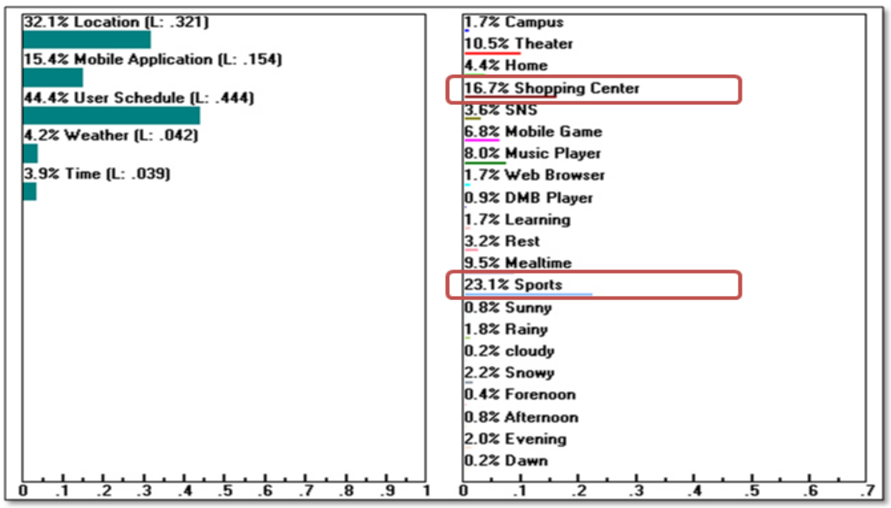


Figure 10. Analysis Result of Relative Importance on Context that Impacts Emotion

This study has matched and indexed emotionally expressive languages on the basis of questionnaire results about emotionally expressive languages that may be felt in a specific context by the users (considering the subjects of questionnaire are university students) and has defined emotionally expressive languages by EARL as described in 5.2. In addition, relative weight can be obtained in each context (not just a single context, but also complex context such as “playing soccer at the playground at college on a cloudy day”) by using AHP. Therefore, it is possible to assess emotion with a method which selects emotionally expressive language with the highest weight value in a complex context.

6. Sentence Generation

On the basis of the results of behavior inference, 4W context will be converted into a short sentence as Table 3 demonstrates. This short sentence will be provided in two types: "behavior" or "movement." For example, if the user is input at Who, morning at When, library at Where, and study at What, a short sentence like "This morning I am studying in a library" will be generated.

Table 3. Rule of Sentence Generation Using 4W1M

Types	4W	Contents of Context	Basic Simple Sentence
Emotion	Emotion	Emotionally Expressive Language	EARL-based Emotionally Expressive Language
Behavior	Who	User	This [When][Who] is doing [What] at[Where]
	When	Time	
	Where	Present location	
	What	Inferred behavior	
On Move	Who	User	At [When] [Who] is one that moves to [Where]

7. Experiments and Evaluation

For the experiment, auto blogging application is implemented with embedded me2API, the open API of me2day (me2day, http://www.me2day.net/) provided by NHN Corp[15]. For system evaluation, data are collected for 20 weeks from 12 participants(divided three groups) for evaluation (an initial 3-week data are used as training data since there is no user feedback). In particular, user feedback is evaluated its correctness of auto blogging information (blogging sentence, reasoned emotional information) provided by the system, with the evaluation result being provided 8 times over the 15 weeks on a working hour basis in a day. The evaluation result of each group is presented in Figure 11 and its correctness is normalized with the calculation of the accumulative mean (last 5 weeks) for each group.

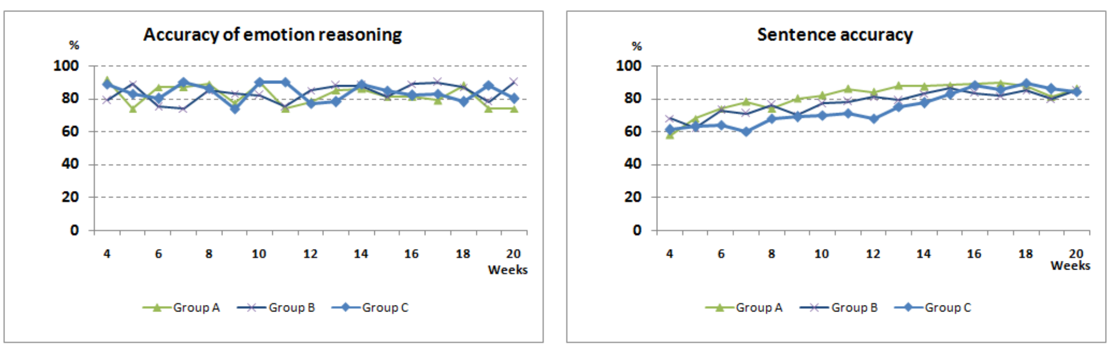


Figure 11. Result of Feedback on Sentence and Emotion's Reasoning Accuracy by Groups

It takes 6 weeks to have sufficient training date for the experiments. The result shows that prediction accuracy increases up to 85.4% in major locations and the accumulative mean for correctness of reasoning emotion information is 82.6%.

8. Conclusion

It is necessary to reason user behavior by using the location and time context of users for the micro auto blogging service by means of users' recognition of context. Location and time context should be managed and analyzed based on a sequence; however, this paper executes a granular tree-based context modeling for the efficient processing of context. In addition, it is verified that user schedule reasoning is possible by using Naive Bayes Classification on the basis of a modeled context. Furthermore, detailed users' behavior of reasoned schedule is possible by using a context tree. Furthermore, this paper has adopted an AHP for users' emotional reasoning to match with detailed behavior and has verified that users' emotional reasoning is possible with higher probability.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation (NRF) funded by the Ministry of Education, Science and Technology (2012042625).

References

[1] A. B. Kocaball and A. Koçyiğit, "Granular Best Match Algorithm for Context-Aware Computing Systems", ACS/IEEE International Conference, **(2006)**, pp. 143-149**.**
[2] Y.Y. Yao, "The Rise of Granular Computing", Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), to appear, **(2008)**.
[3] Z. Qin, "Naive Bayes Classification Given Probability Estimation Trees", ICMLA'06 5th International Conference, **(2006)**, pp. 34-42.
[4] P. Korpipaa, M. Koskinen, J. Peltola, S. -M. Makela and T. Seppanen, "Bayesian approach to sensor-based context awareness", Personal and Ubiquitous Computing, vol. 7**, (2003)**, pp. 113-124.
[5] T. L. Saaty, "The Analytical Hierarchy Process", Mc Graw Hill, New York, **(2008)**.
[6] Statistics Canada, General social survey on time user, http://www.statcan.ca
[7] S. S. Lim and S. B. Cho, "User Adaptive Schedule Management Agent using Bayesian Networks", Korea Computer Congress, vol. 33, no. 1(B), **(2006)**, pp. 103-105.
[8] H. T. Kim, J. K. Min and S. B. Cho, "Bayesian Inferrence and Context-Tree Matching Method for Intelligent Services in a Mobile Environment", Journal of Korean Institute of Information Scientists and Engineers, vol. 36, no. 2, **(2009)**, pp. 144-152.
[9] HUMAINE Emotion Annotation and Representation Language (EARL), http://emotion-research.net/.
[10] R. Plutchik, "Circumplex Models of Personality and Emotions", Washington, DC: American Psychological Association, **(1997)**.
[11] P. Ekman, "Basic Emotions", The Handbook of Cognition and Emotion, Sussex, U.K.: John Wiley & Sons, Ltd. **(1999)**, pp. 45-60.
[12] D. L. Robinson, "Brain function, mental experience and personality", The Netherlands Journal of Psychology, vol. 64, **(2009)**, pp. 152-167.
[13] W. Parrott, "Emotions in Social Psychology", Psychology Press, Philadelphia, **(2001)**.
[14] Expert Choice, http://expertchoice.com/.
[15] me2day, NHN Corp., http://www.me2day.net/.

Authors

Il-Kyoung Kwon, received the B.S., M.S. degrees in computer science from Kongju National University, Gongju, Korea in 1999, 2001, respectively. His research interests include intelligence systems, emotion modeling and context awareness based personalization services.

Sang-Yong Lee, received the M.S. degree in system science from Tokyo Institute of Technology, Tokyo, Japan in 1988 and the Ph.D. degree in computer engineering from Chung-Ang University, Seoul, Korea in 1993. From 1988 to 1989 he had been a researcher at NEC, Tokyo, Japan. Also he had been a visiting professor at University of Central Florida, Orlando, USA from 1996 to 1997. Since 1993 he has been a professor of the division of computer science and engineering, Kongju National University, Gongju, Korea. His research interests include intelligent systems, context awareness and ubiquitous computing.