A Fuzzy Logic System to analyze a Student's Lifestyle

Sourish Ghosh*, Aaditya S. Boob*, Nishant Nikhil*, V. Nayan Raju*, Ankit Kumar*, and S. K. Barai†
*Department of Mathematics
Email: {sourishg, aaditya733, nishantnikhil, v.nayanraju, ankitkahnani}@iitkgp.ac.in
†Professor, Department of Civil Engineering
Email: skbarai@civil.iitkgp.ernet.in
Indian Institute of Technology, Kharagpur

***Abstract*—A college student's life can be primarily categorized into domains such as education, health, social and other activities which may include daily chores and travelling time. Time management is crucial for every student. A self realisation of one's daily time expenditure in various domains is therefore essential to maximize one's effective output. In this paper we present how an Android application using Fuzzy Logic and Global Positioning System (GPS) analyzes a student's lifestyle and provides recommendations and suggestions based on the results.**

Keywords—Fuzzy Logic, GPS, Android Application

I. INTRODUCTION

A college student's life is multidimensional. Students are expected to be academically excellent, physically fit and socially active along with managing their daily chores and pursuing their fields of interest. This structure would not only help students to engage all activities but also help them live a balanced life. This practice would eventually help them make better career choices on the basis of their interests. For such a practice one needs to invest a threshold amount of time and effort in all the activities. However only a certain amount of students are involved and excel in such a practice. This paper discusses a novel approach using fuzzy logic to generate an analysis of a student's daily time expenditure in these various categories. Based upon the analysis of the results obtained from the above data appropriate results must be provided on regular basis. This would help the students work in their non-performing fields and maintain a balanced lifestyle.

A. About Fuzzy Logic

Over the past three decades, fuzzy logic is widely used in all problem-solving domains. One of the reasons for such instantaneous growth since its inception is its usability across all sectors be it Dynamic Programming, Process Control or Optimization. Fuzzy logic discards the theory of 'Absolute Truth'and instead proposes a new theory of 'Partial Truth', also referred as degree of membership (suggested by Prof. Zadeh in 1965).

Let $\mathbf{S}$ be a non empty set, called the *universe set*. Now, consider any crisp set $A \subset \mathbf{S}$. A characteristic function $\boldsymbol{\chi}_A$ is defined as

$$\boldsymbol{\chi}_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$$

A characteristic function assigns value of either 0 or 1 to each element of $\mathbf{S}$. Now consider a fuzzy set $B \subset \mathbf{S}$. A membership function $\boldsymbol{\mu}_B(x)$ is defined as $\boldsymbol{\mu}_B : \mathbf{S} \to [0, 1]$. Unlike the notion of a set in classical set theory where an element either belongs or does not belong to a particular set based on a bivalent condition, in fuzzy set theory an element's belongingness to a particular set is decided using membership function which gives a membership value between 0 and 1.

B. Problem Formulation

The problem can be divided into three major parts:

- **Data Collection**: Using GPS, we first collect where and how long the user spends his/her time and tabulate that data. For this we use the Google Places API.
- **Fuzzification**: Fuzzify the crisp input and calculate the values of corresponding membership functions.
- **Defuzzification**: Set up a fuzzy inference system based on certain rules and then return recommendations and suggestions.

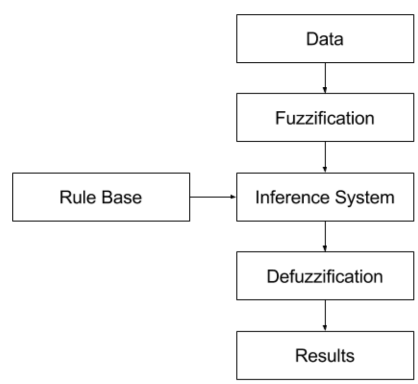

Fig. 1: Fuzzy Logic System

II. WORKING PRINCIPLE

A. Data Collection

A college student is carrying his/her smart phone everywhere. Hence using the GPS we can extract his/her position throughout the day. In the application and testing of this paper, the mobile application was developed on Android while the point of interest was extracted using Google Maps API by querying the user's location extracted from GPS. Google Maps

TABLE I: Sample locations and purposes

Location	Purpose
Cafe, Restaurant	Going out with friends and family.
Supermarket, Gas Station	Chores
Gym, Ground, Hospital	Exercise or Health Treatment
Cinema Hall, Spa	Leisure and Relaxation
Bank, Business Associates	Work

API classifies most of the locations into various categories namely `restaurant`, `shopping_mall`, `city_hall` etc. Let us refer to all these categories hence forward as *tags*. Apart from these existing tags, we generate two additional tags namely `home` and `work`. The GPS data for these two additional tags would be user specific. Hence initially every user needs to update their location for these two tags specifically. This step is conducted so as to recognize distinctly one's home and workplace which in further course would generate accurate results. Let us consider this example, one might go to a pizza shop to hang out with friends and family. However if someone is working in a pizza shop and the GPS details of the specific pizza shop is not known beforehand, it is very likely that one might consider this entire working time as time utilized for hanging out with friends. However if the person goes to some other pizza shop it is very likely he is going out with friends. To avoid this confusion this initial step has to be carried out.

Let $\mathbf{X}$ be the set of all tags defined as $\mathbf{X} = \{x \mid x \text{ is a tag}\}$. Analysing the way a person lives is governed by many parameters, but in a typical student's life we are mainly concerned about one's health, education, leisure and social life. However a person also invests certain amount of time which fails to fall under these categories. A example of this would be travelling time. Activities like these fall under the *other* category. Now let S, L, H, W, O be subsets of $\mathbf{X}$ defined as

$$
\begin{aligned}
S &= \{x \mid x \in \mathbf{X},\ x = \text{social and } x \neq \text{home, work}\} \\
L &= \{x \mid x \in \mathbf{X},\ x = \text{leisure and } x \neq \text{home, work}\} \\
H &= \{x \mid x \in \mathbf{X},\ x = \text{health and } x \neq \text{home, work}\} \\
W &= \{x \mid x \in \mathbf{X},\ x = \text{work and } x \neq \text{home}\} \\
O &= \{x \mid x \in \mathbf{X} \text{ and } x \notin S \cup L \cup H \cup W\}
\end{aligned}
$$

A *tag* x might belong to one or more of the sets S, L, H, W. For example, a person might visit an Amusement Park. In this case the person's *social* and *leisure* scores are both incremented. Using this categorization technique we can extract one's location and time spent at each *tag* for the entire day. TABLE I lists down some locations and their possible purpose of visits. The locations mentioned are basically *tags* other than `home` and `work`.

Weighing criteria: For a given purpose, different locations would have different amount of productivity and impact. For example, hospital and gym both fall under the *health* category. However one visits a gym to increase his physical activity and hence visiting a gym has a positive impact on one's health. However one visits a hospital if he/she has fallen sick. Hence, visiting a hospital has a negative impact on one's health. So we have to handle these two situations differently.

We define a function $Y : \mathbf{X} \to \mathbb{R}$ such that $Y(x)$ for every $x \in \mathbf{X}$ denotes the time spent at the location *tag* x. For example, let $x =$ gym. Say $Y(x) = 0.5$. This implies a person has spent 30 minutes at a *gym* in the entire day. The unit of time is set in hours throughout this paper.

We define a function $Z_S : S \to [-100, 100]$ such that $Z_S(x)$ for every $x \in S$ denotes the intensity of the *tag* x with respect to the *social* category. Similarly we define Z_H, Z_L, Z_W, and Z_O for the *health*, *leisure*, *work*, and *other* categories respectively. The range $[-100, 100]$ is chosen for normalization purposes. For example, let $x =$ gym. Say $Z_H(x) = 50 > 0$ as gym has a positive health impact. Let $y =$ hospital, then $Z_H(y) = -20 < 0$ as hospital has a negative health impact. However, $Z_L(x) = Z_L(y) = 0$ as both x and y don't contribute to the *leisure* category. Also note that if a *tag* t belongs to two different categories, then its weightage in both the categories cannot be 0.

For both Y and Z we have excluded the *home* tag as it is a special case. This is explained later.

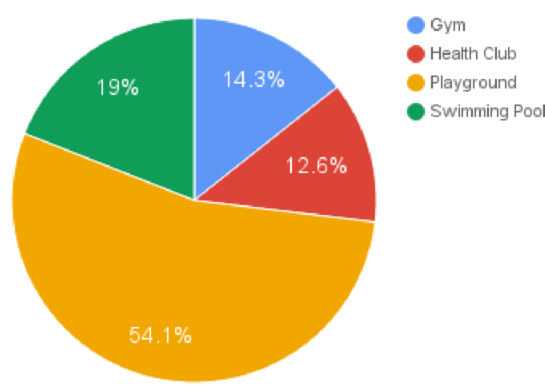

Fig. 2: Survey for *positive* health weights

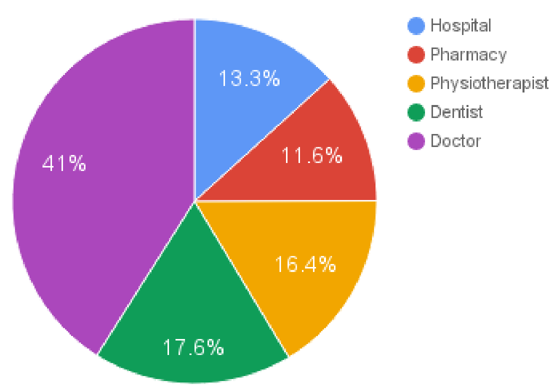

Fig. 3: Survey for *negative* health weights

Assigning weights: One is free to assign the weights independently. However for better results, we assign weights by conducting a survey to understand how appropriately a location *tag* fulfils the purpose of a category. For instance consider the *health* category. In the survey we ask a sample population to rank every $x \in H$ in an order of fulfilment of their positive *health* benefits. Consider the following survey with

$$H = \{\text{gym}, \text{ playground}, \text{ swimming\_pool}, \text{health\_club}, \\ \text{hospital}, \text{ pharmacy}, \text{ physiotherapist}, \text{ dentist}, \text{ doctor}\}$$

Fig. 2 shows a survey for determining *positive* weights in the *health* category. As 54.1% people taking the survey voted *playground* as their maximum positive benefit from the *health* category, the corresponding weight for $x = \text{playground}$ is computed as $Z_H(x) = \frac{54.1}{100} \times 100 = 54.1$.

Fig. 3 shows a survey for determining *negative* weights in the *health* category. As 41% people taking the survey voted *doctor* as their maximum non fulfilment from the *health* category, the corresponding weight for $x = \text{doctor}$ is computed as $Z_H(x) = -\frac{41}{100} \times 100 = -41$.

Home tag: The time spent at the *home* location might not be entirely used for rest and leisure purpose only. One might practice yoga at one's home and the equivalent time should be added to the *health* category. Let τ denote the total time spent at *home*. And τ_H, τ_W, τ_L, τ_O, τ_S denote the equivalent time in respective categories. This time is taken as user input through the mobile application. For better results a random push notification system can be used to learn the characteristics of the user. The home tag will be associated with weights ξ_H, ξ_S, ξ_L, ξ_W, ξ_O which denote the intensity of the tags at *home*. For instance, $\xi_W = 30$ and $Z_W(\text{office}) = 50 > 30$ as working at *home* might not be as productive as working at *office*.

B. Fuzzification

Fuzzification of time: Consider a person p. Suppose p visits tags $\{x_1, x_2, \ldots, x_n\}$, with the time spent at these locations denoted by $\{Y(x_1), Y(x_2), \ldots, Y(x_n)\}$. Let K_H, K_L, K_S, K_W, K_O denote the overall time spent in *health*, *leisure*, *social*, *work*, and *other* categories respectively. Then

$$K_H = \sum_{x_i \in H} Y(x_i) + \tau_H$$

Similarly, K_W, K_S, K_L, K_O are defined.

We define the following fuzzy sets for all the categories. These sets define the type of lifestyle a person is living in each category. Here *leisure* also includes rest.

$$\begin{aligned} \textit{health} &= \{\text{unfit}, \text{ fit}, \text{ proactive}\} \\ \textit{leisure} &= \{\text{hectic}, \text{ ideal}, \text{ lazy}\} \\ \textit{social} &= \{\text{reserved}, \text{ sociable}, \text{ over\_social}\} \\ \textit{work} &= \{\text{lethargic}, \text{ hard\_working}, \text{ industrious}\} \\ \textit{others} &= \{\text{non\_productive}, \text{ productive}\} \end{aligned}$$

The membership functions for these fuzzy sets are constructed by conducting a survey on a sample population. We will approximate the data from the survey using quantile range and trapezoidal membership functions. However, one can use various other techniques to plot membership functions. For instance, in a sample survey the hours spent by fit students in the *health* category were: 0.45, 1.25, 2, 2.25, 2.5, 2.5, 2.75, 2.75, 3, 4, 4.25. So with respect to the inter quantile range $Q_1 = 2$, $Q_2 = 2.5$, $Q_3 = 3$, $\inf = 0.45$, and $\sup = 4.25$. The trapezoidal membership function for the linguistic term "fit" under the *health* category using these values is shown in Fig. 4.

Figure 5, 6, 7, 8, and 9 show the membership functions for each linguistic term under each category.

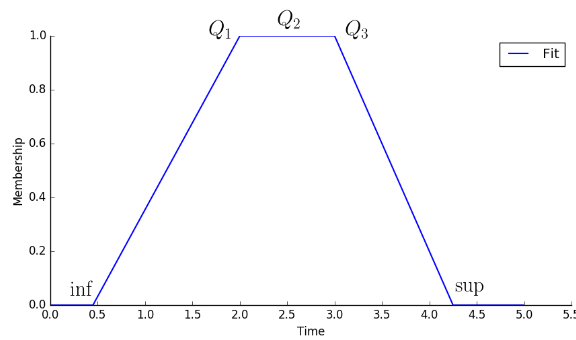

Fig. 4: Membership function for the linguistic term *fit*

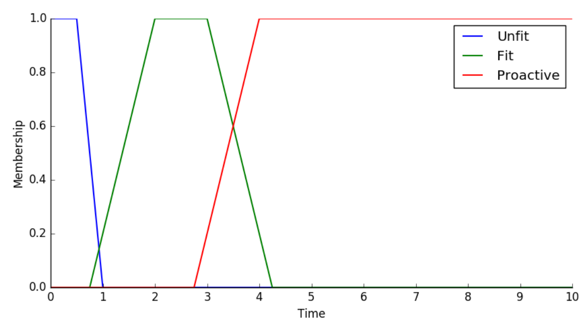

Fig. 5: Membership function for *health* category

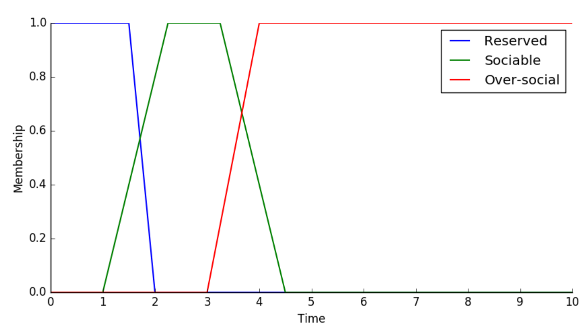

Fig. 6: Membership function for *social* category

Fuzzification of score: Not only the time spent at a location is important but also how the time is spent is important too. This effective utilisation of time is denoted by a *score* M_S, M_L, M_O, M_W, and M_H for the respective categories.

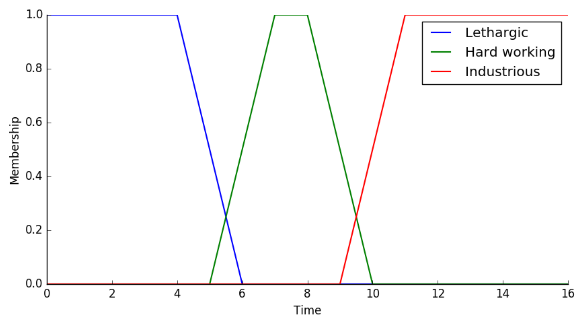

Fig. 7: Membership function for *work* category

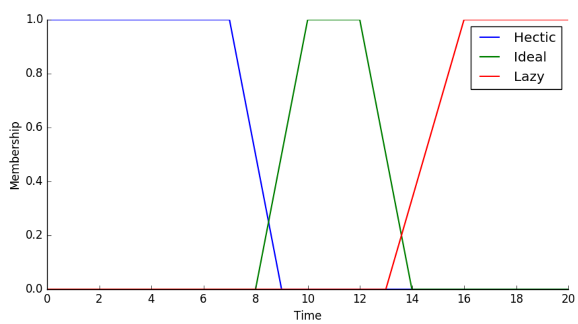

Fig. 8: Membership function for *leisure* category

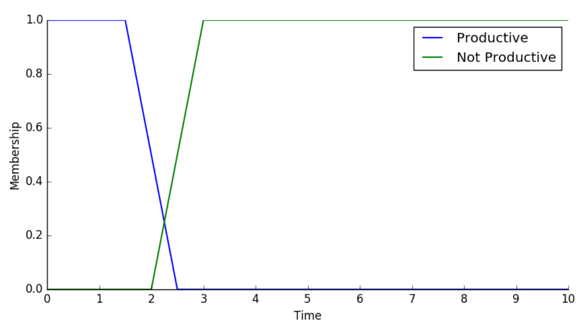

Fig. 9: Membership function for *other* category

The score for the *social* category is calculated as follows

$$M_S = \sum_{x \in S} Y(x) Z_S(x) + \tau_S \xi_S$$

Similarly we define the other scores. The fuzzy set of linguistic terms "low_score", "ideal_score" and "high_score" define the fuzzy scores in each category. The membership function of these sets in all categories is calculated similar to the fuzzy time membership functions by conducting a survey. For instance, a survey conducted on a sample of fit students is shown in TABLE II. Hence $\inf = 11.25$, $Q_1 = 29.75$, $Q_3 = 42$, $\sup = 50$. Accordingly the membership function for the linguistic term "ideal_score" under the health category is shown in Fig. 10. Similarly we can plot plot the membership functions for the entire fuzzy set across all categories.

TABLE II: Survey to determine the membership function for the linguistic term "ideal_score" under *health* category

Time (t_i)	**Weight** (w_i)	**Score** $(\sum_i w_i t_i)$
0.45	25	$0.45 \times 25 = 11.25$
$1,\ 0.25$	$10,\ 12$	$1 \times 10 + 0.25 \times 12 = 13$
2	15	$2 \times 15 = 30$
2.25	18	$2.25 \times 18 = 40.5$
2.5	18	$2.5 \times 18 = 45$
2.5	20	$2.5 \times 20 = 50$
2.75	12	$2.75 \times 12 = 33$
$2,\ 0.75$	$13,\ 10$	$2 \times 13 + 0.75 \times 10 = 33.5$
3	11	$3 \times 11 = 33$
$2,\ 2$	$13,\ 8$	$2 \times 13 + 2 \times 8 = 42$
4.25	7	$4.25 \times 7 = 29.75$

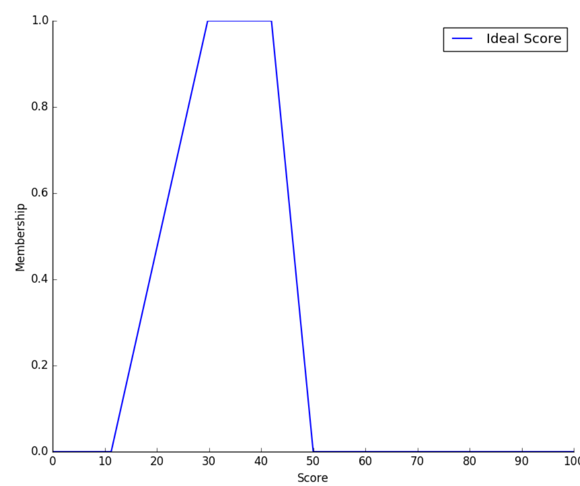

Fig. 10: Membership function for "ideal_score" under *health* category

C. Defuzzification

Given the input data, $Y(x)$, $Z_i(x)$, τ_i, and ξ_i where $x \in \mathbf{X}$ and $i = S,\ L,\ O,\ W,\ H$, we calculate corresponding K_i and M_i. Using surveys we determine the membership functions for all linguistic terms in all categories for both fuzzification of time and score. Then we determine the membership value of K_i and M_i in the respective categories for all the linguistic terms. Let $\{R_1, R_2, \ldots, R_N\}$ be a set of recommendations. Now every R_k $(1 \leq k \leq N)$ will be dependent on a set of linguistic terms. For example, a recommendation $R =$ *"All work and no play makes Jack a dull boy."* will be outputted if a person is spending too much time and effort in work and less in his leisure and social life. That is he/she has

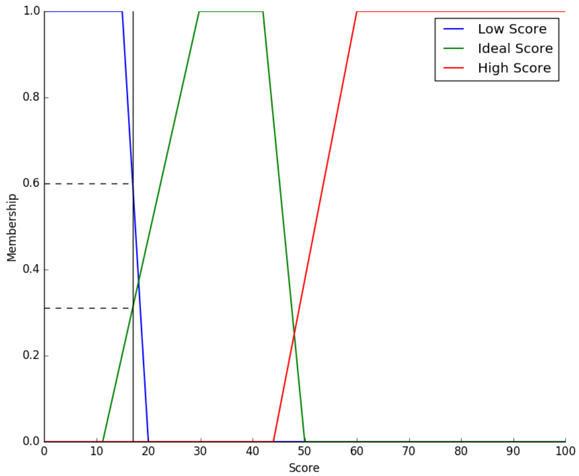

Fig. 11: Calculation of membership values

a "industrious" work life with a high work score and has a "reserved" social life with a low social score and a "hectic" life with respect to leisure with a low score. So attributes of R can be represented as $\{K_W$ = "industrous", M_W = "high_score", K_S = "reserved", M_S = "low_score", K_L = "hectic", M_L = "low_score"$\}$.

Let R_k be a recommendation with attributes $\{a_1, a_2, a_3, \ldots, a_n\}$. Here each a_j $(1 \leq j \leq n)$ is a combination of score/time with respect to a linguistic term of a category. Hence as shown previously we can calculate its membership value. Let $\mu_1, \mu_2, \mu_3, \ldots, \mu_n$ denote the respective membership values for each attribute. Here n can vary for each R_k. For instance, Fig. 11 shows the membership functions of M_H. Let a_1, a_2, a_3 be the following attributes

$$\begin{aligned} a_1 &= M_H : low\_score \\ a_2 &= M_H : ideal\_score \\ a_3 &= M_H : high\_score \end{aligned}$$

Hence μ_1, μ_2, μ_3 for $M_H = 17$ as seen from Fig. 11 will be 0.6, 0.310, 0.0 respectively. Using equal weighing criteria for each a_j, we can calculate a score of each recommendation $\rho(R_k)$ defined as

$$\rho(R_k) = \frac{1}{n} \sum_{j=1}^{n} \mu_j$$

Now, using the *most probable criterion* the recommendation with the maximum score value $\rho(R_k)$ will be displayed as output.

III. Experiment

A survey conducted in IIT Kharagpur was conducted to determine all the membership functions for all linguistic terms across all the categories. Some of the membership functions are shown in this paper. We also installed the mobile application on students' smart phones and analysed the results. We picked a random student and analysed his data for a day. TABLE III shows the *tags* he visited throughout the day and their

TABLE III: Experiment: Data

Tag	**Time**	**Weight**	**Score**
university	$Y(x) = 6$	$Z_W(x) = 50$	300
library	$Y(x) = 4$	$Z_W(x) = 20$	80
home	$\tau_W = 2$	$\xi_W = 30$	60
	$\tau_S = 0.5$	$\xi_S = 30$	15
	$\tau_H = 0.5$	$\xi_H = 20$	10
	$\tau_L = 6.5$	$\xi_S = 30$	195
	$\tau_O = 1$	$\xi_O = 10$	10
cafe	$Y(x) = 1$	$Z_S(x) = 20$	20
supermarket	$Y(x) = 1$	$Z_O(x) = 9$	9
grocery	$Y(x) = 0.5$	$Z_O(x) = 10$	5
travel	$Y(x) = 1$	$Z_O(x) = 15$	15

TABLE IV: Experiment: Calculation of total time and score for each category

Total time	**Total score**
$K_S = 1.5$	$M_S = 35$
$K_L = 6.5$	$M_L = 195$
$K_O = 3.5$	$M_O = 39$
$K_W = 12$	$M_W = 440$
$K_H = 0.5$	$M_H = 10$

corresponding time and weights. The score for each *tag* is also enumerated. TABLE IV shows the total time and score across all the categories. We considered the recommendations in the set R where $R = \{R_1, R_2, R_3, R_4\}$.

R_1 = *"Catch up a movie this evening."*
R_2 = *"Work is worship."*
R_3 = *"Family matters."*
R_4 = *"Hit the gym."*

The attributes of R is shown in TABLE V. The membership values for each attribute is shown in TABLE VI and the corresponding score of each recommendation is also enumerated. As $\rho(R_1)$ is maximum the mobile application would recommend the student to *"Catch up a movie this evening."*

IV. Conclusion

Time management has always been a difficult art to master. This paper helps one master it by using fuzzy logic to understand the science behind this art. The use of this technology in the long run would lead to more accurate results. The main highlights of this method is that most of the segments are self

TABLE V: Experiment: Recommendation attributes

Recommendation	**Attributes**
R_1	$\{K_L$ = "hectic", M_L = "less_score", K_W = "industrious", M_W = "high_score"$\}$
R_2	$\{K_W$ = "lethargic", M_W = "less_score", K_L = "lazy", M_L = "high_score"$\}$
R_3	$\{K_S$ = "reserved", M_S = "less_score"$\}$
R_4	$\{K_H$ = "unfit", M_H = "less_score"$\}$

TABLE VI: Experiment: Recommendation score calculation

Recommendation	**Membership Values** (μ_j)	**Score** $(\rho(R_k))$
R_1	$\{1.0,\ 0.8,\ 1.0,\ 1.0\}$	0.95
R_2	$\{0.0,\ 0.0,\ 0.0,\ 0.0\}$	0.0
R_3	$\{1.0,\ 0.7\}$	0.85
R_4	$\{1.0,\ 0.8\}$	0.9

configurable. The dynamics of a student studying in university A and that of a student in university B can be very different due to a lot of factors such as geographic location, nature of university and infrastructure. Hence the parameters for analysis of these two students should be significantly different. This paper gives the user this flexibility to configure various attributes such as weight of tags, type of membership functions and the set of recommendations. This practice would yield better results. The long term applications are vast. This method can be used to access the performance of all students in an university. The results can be shared with the university so that the university can take appropriate actions for the welfare of the students in general.

ACKNOWLEDGMENT

The authors would like to thank Professor Sudhir Kumar Barai of the Civil Engineering Department, IIT Kharagpur for his continued and unconditional guidance. An exemplary teacher and a magnificent person, we consider ourselves lucky to have been taught by him and to have worked under his supervision. Without his course, "Soft Computing Tools in Engineering", and his co-operation the preparation of this paper would not have been possible.

REFERENCES

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.