



Measuring Educational Attainment: A Proposed Fuzzy Methodology

Indrajit Ghosh

Artificial Intelligence Laboratory, Dept. of Computer Science,
Ananda Chandra College, Jalpaiguri, West Bengal, India

Abstract: *For measurement of human development, literacy rate is considered to be an important indicator. But this indicator may not be a good one to capture the qualitative dimension to an individual's or society's educational attainment. The overall educational level of an area is an important issue beyond the literacy rate. The overall educational level can be thought of as an outcome of the educational levels of individuals. But there is no well defined algorithm and mathematical model available to detect the overall educational level of an area. A heuristic approach based on accumulated experience of an expert is considered to be an effective approach. As fuzzy logic offers a natural and convenient framework in modeling various concepts in social science domain, this paper presents the implementation of fuzzy logic to develop a mathematical model for detection of overall educational level of an area in terms of Education Index. The contribution of the study is two folds: conceptualization of "Education Profile" and proposing a new mathematical model to measure educational attainment in terms of "Education Index".*

Keywords: *Educational attainment, Education Profile, Education Index, Fuzzy system.*

I. INTRODUCTION

The process of education and attainments thereof has been an impact on all aspects of life. A number of attempts to measure educational attainment (across countries) have been made. The earlier empirical studies used literacy rate or school enrolment ratio. The UNDP in its HRD Report, 1990, pointed out that literacy is a person's first step in learning and knowledge building and thereof, literacy indicator was essential for measurement of human development [1]. But literacy is a partial proxy for education and focuses on a particular aspect of the education and to that extent, captures only a limited dimension of educational attainments.

Traditionally, the society's literacy is measured by the literacy rate which is the mere percent of the adult population that is literate. But the adult literacy rate may measure only a superficial capacity to read and write one's name or a simple sentence and hence, may not be a good indicator in itself for capturing educational attainment of a society, particularly when it is a result of mass adult literacy program and not an outcome of a formal education system. Moreover, a person having eighth standard can not have the same contribution as a graduate or a post graduate in educational attainment but both of them have the equal contribution in measuring the literacy rate. So mere literacy rate may not be a good indicator if one is looking at the qualitative dimension to an individual's or a society's educational attainment [1].

Though some important works on literacy have been reported [2-10], but they are mainly concerned with the intra-household, isolated and proximate illiteracy arising from the traditional literacy concept based on classical mathematics. But none of them focuses on the qualitative dimension of educational level or status.

A lot of works have been published on cross-nationally comparable measurement of educational attainment which includes [11-16], instead of absolute one. In comparable research, two schemes have been proposed. One is Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) [17] and the other is International Standard Classification of Education 1997 (ISCED-97) [18,19]. But both the schemes use the categories with sharp boundaries. Still, the problem with using CASMIN for comparative research is that it has only been applied to a limited number of countries and there is no formal guidelines for its implementation in other countries [20]. So a well developed conceptual model to measure educational attainment is needed to be proposed.

The notion of fuzzy logic was conceptualized by Lotfi Zadeh in 1965 [21]. It offers researches an interpretive algebra, a language that is half-verbal-conceptual and half-mathematical-analytical [22]. As a humanistic science, economics should thus have been one of the early prime targets for utilizing fuzzy logic. Claude Ponsard was the pioneer to initiate the formulation of economic theory by taking advantage of fuzzy set theory [23,24]. Some important contributions of some eminent economists are overviewed in a special issue of *Fuzzy Sets and Systems* [25]. A paper by Chen, Lee and Yu [26] is a readable overview of some issues of fuzzy economics. The principal source of information regarding the current states of the use of fuzzy logic in economics is a book by A. Billot [27]. Other relevant works regarding the application of fuzzy set theory in economics include Cerioli and Zani [28], and Martinetti [29]. But none of them are concerned to find out the indicators for measuring educational attainment.

In 2001, National Human Development Report pointed out that there could be alternative indicators or combination of indicators and consequently some methodologies that could be suggested to reflect some or most of the considerations in

capturing the educational attainment in the country [1]. Beyond the literacy rate, consideration of other alternative one should be an important issue.

Instead of mere literacy rate, the educational attainment can be thought of as an outcome of the ‘*Education Profile*’ (percentages of population having different educational level of individuals). But there are still no well defined mathematical models and methodologies available to measure the educational attainment of an area or country, based on the education profile. This encourages for an extensive study and promotes to propose a conceptual mathematical model to measure the educational attainment.

This work attempts to reformulate an alternative index, Education Index (EI) for measuring educational attainment in the light of most promising fuzzy logic proposed by Prof. L. Zedah. “Education Index” (EI), ranging from 0 to 100, as an outcome of the education profile, might be considered as a graceful alternative to measure the educational attainment with qualitative dimension.

The next section presents an overview of fuzzy logic system. Section III illustrates the concept of Education Profile and how fuzzy logic is implemented. Section IV contains a case example to present how this model can be used to measure the Education Index, for better understanding. At last, conclusion is provided in section V.

II. AN OVERVIEW OF FUZZY LOGIC SYSTEM

The concept of fuzzy logic was first introduced in 1965 by Prof. Lotfi Zadeh [21] to account for the real world gradient that exist between true and false (yes and no). It deals with degree of membership and degrees of truth instead of yes and no. Fuzzy logic allows the programmer to deal with natural ‘linguistic sets’ of states such as low, high, medium etc. Classically, a set is defined by its members. An object may be either a member or a non-member. The connected logical propositions may be true or false. This concept of crisp set may be extended to fuzzy set with the introduction of the idea of partial truth. Any object may be a member of a set “to some degree”; and a logical proposition may hold true to some degree. The assignment of Degree of Membership (DOM) to a variable is done through Membership Functions (MFs). A membership function can be of bell-shaped, s-shaped, reverse s-shaped, triangular or trapezoidal. These membership functions are defined for each input and output. The process of assigning membership functions to an input or output linguistic values is called the *fuzzification* process.

Any fuzzy usually takes multiple real-world inputs, fuzzify these inputs and produce a single real-world output. After fuzzy inputs are determined, the system goes through rule evaluation. Rule evaluation involves a series of if-then statements. The process of determining the context to which each rule is relevant to the current set of inputs and then determining the conclusion from the inputs and the rule base is collectively called *inference*.

There may be multiple rules and after each rule is evaluated, it can be defuzzified and produce a real world output. Defuzzification combines the results of each rule into a single unique result in terms of crisp value. It is the process to convert the output of the fuzzy process logic into ‘crisp’ – numerically precise solution variables.

III. MATHEMATICAL FOUNDATION

A. Education profile

This method has been developed to compute the “*Education Index*” as a measure of educational attainment based on education profile. Education profile is a profile of percentages of population of different educational level of individuals. The range of educational level of individuals have been considered to be divided into five categories as shown in Table I. Here the categories of educational level provided by UNESCO are not used because it does not consist any categories beyond post-secondary level.

TABLE I

<i>Category</i>	<i>Educational level</i>
Category-1	Illiterate
Category-2	5 years of schooling
Category-3	12 years of schooling
Category-4	Bachelor degree
Category-5	PG/ Tech./ and above

Now the education profile that has been considered in this model consists of the percentages of the population for Category-1 (POC₁), Category-2 (POC₂), Category-3 (POC₃), Category-4 (POC₄) and Category-5 (POC₅) respectively. Let us consider an example for illustration. Suppose for an area (Chapani village of Alipurduar Block-II, West Bengal, India), the percentage of population for different Categories obtained from direct field survey by the author is presented in Table II.

TABLE II

Educational level	Variables	Percentage of population
Category-1	POC ₁	14
Category-2	POC ₂	41

Category-3	POC ₃	23
Category-4	POC ₄	13
Category-5	POC ₅	9

The Education Profile is $\epsilon_P = \{ POC_1, POC_2, POC_3, POC_4, POC_5 \} = \{ 14, 41, 23, 13, 9 \}$.

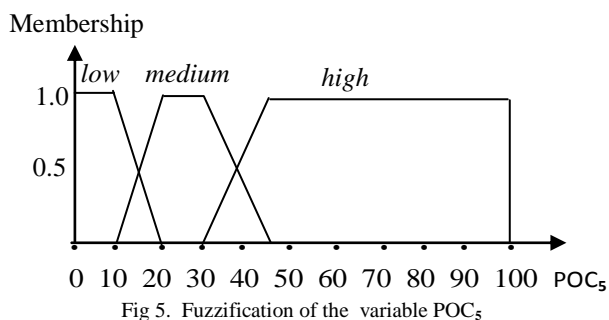
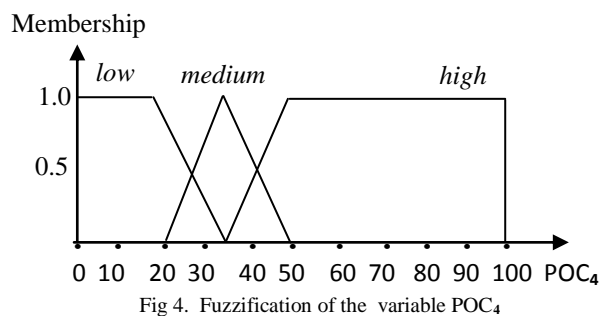
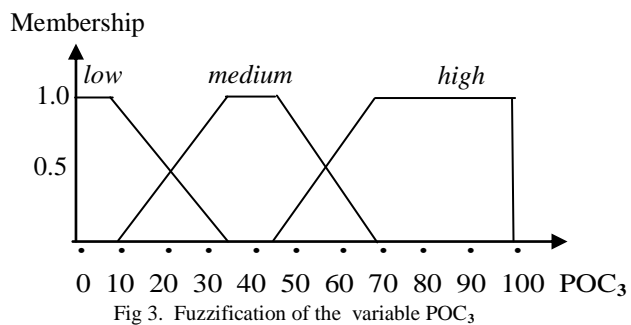
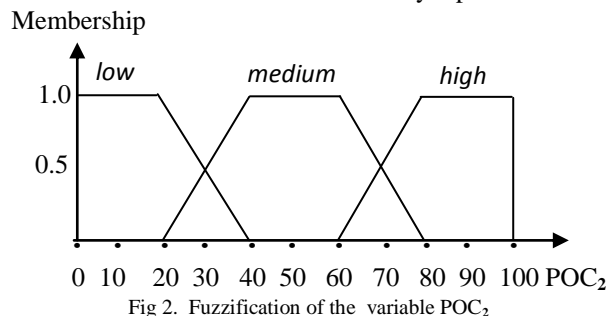
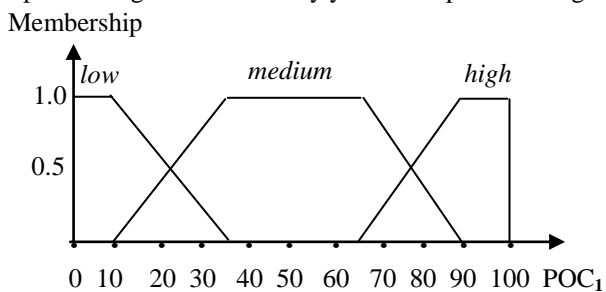
Since the input of the system is the education profile of an area, so the input variables to the fuzzy model are the percentages of population having different levels such as POC₁, POC₂, POC₃, POC₄ and POC₅ respectively. These variables are treated as the state fuzzy variables. The output of the fuzzy model is the education index which varies in accordance with the variation of the state fuzzy variables.

B. Fuzzification of system input variables

Fuzzification involves translation of propositions into quantitative values using membership functions. The first step in the development of the fuzzy logic based method is to construct fuzzy sets for the input and output parameters for various linguistic variables such as low, medium, high etc.

In fuzzy systems, the physical indicators are identified as fuzzy variables. A fuzzy variable describes the discrete values of the fuzzy variables with linguistic commonsense terms. These linguistic terms are called fuzzy-set values. For example, the variable for the percentage of illiterate person (POC₁) is a fuzzy variable that takes *low*, *medium* and *high* as fuzzy-set values. Different fuzzy quantification may be done depending on the number of fuzzy-set values (fuzzy-regions).

As the education profile consists of five variables, POC₁, POC₂, POC₃, POC₄ and POC₅, so they have been considered as the fuzzy input variables. Each of these fuzzy variables is decomposed into three fuzzy regions (*low*, *medium* and *high*) following the rules of thumb [30]. These fuzzy sets are designed based on the accumulated knowledge of the domain expert having more than thirty years of experience. Figures 1 to 5 show the fuzzification of the five fuzzy input variables.



C. Fuzzification of system output variables

The output of this model is the **Education Index** (EI). The education profile of different areas vary in accordance with the variations of the input variables that may change from area to area. For instance, the educational attainment of an area with high percentage of technical diploma or higher degree (category-5) holders is usually higher than that of with high percentage of illiterate persons (category-1).

For better precision, the fuzzy output variable education index (EI), ranging from 0 to 100, is decomposed into five fuzzy regions, *very low*, *low*, *medium*, *high* and *very high* as shown in the Figure 6.

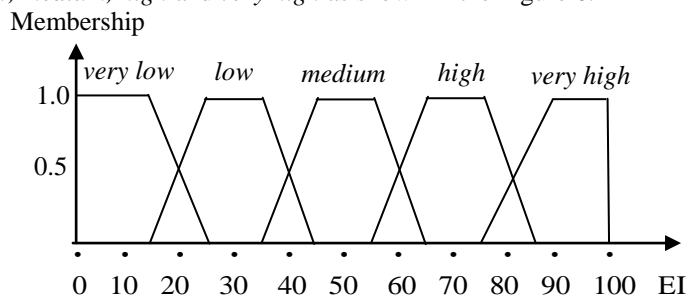


Fig 6. Fuzzification of the output variable EI

Fuzzification of variables lies under the trade-off of precision in detection and computation time and memory. Increasing the number of fuzzy regions would improve the precision of detection but on the other hand, would eventually lead to an increase in the number of decision-making rules and hence the system complexity.

D. Fuzzy inference

The fuzzy input and output variables are related to each other with a rule-based system. A statement about the resulting output variable has to be made for all possible combinations of the categories of all the variables. The process of determining the extent to which each rule is relevant to the current set of inputs and then determining the conclusion from the inputs and the rule base is collectively called inference.

The system is designed to aid the decision-making process for detection of educational attainment of an area based on the education profile. Once the fuzzy input variables are defined, the knowledge representation using rules become easy. The fuzzy production rules of inference are used to compute the education index.

The reasoning is performed by forward chaining with its fuzzy rule-based knowledge. "IF { F } THEN { R }" rules have been used for inferences where {F} is the set of findings, {R} is the detected education index. The format chosen for the definition of rules allows flexibility in structuring the knowledge. An antecedent of any rule may be a composite of number of clauses connected via the logical operations .AND. and .OR.

Examples of a decision making rules are:

If POC₁ is *high* AND POC₂ is *medium* AND POC₃ is *low*
AND POC₄ is *low* AND POC₅ is *low*
THEN EI is *low*.

If POC₁ is *low* AND POC₂ is *medium* AND POC₃ is *high*
AND POC₄ is *medium* AND POC₅ is *low*
THEN EI is *medium*.

Different rules may be activated at the same time and combination of their outputs is then defuzzified to compute the education index.

E. Defuzzification

This final step of the fuzzy inference system is necessary for combining the results of each rule into a single unique result. The output fuzzy labels (*very low*, *low*, *medium*, *high* *very high*) are used to produce the final system output. A fuzzy output label has a corresponding output value. When each rule is evaluated, the minimum numeric value of its antecedent is taken to be the rule strength. However, when determining the output fuzzy label, the maximum numeric value is taken to be the label's value.

IV. COMPUTATION OF EDUCATION INDEX: A CASE ILLUSTRATION

Let us consider the education profile of an area, as mentioned in section III.A, which has been taken as a case study for the proposed model. The education profile obtained is $\epsilon_P = \{ POC_1, POC_2, POC_3, POC_4, POC_5 \} = \{ 14, 41, 23, 13, 9 \}$. Now the first input is POC₁ = 14 which has the membership values (see Figure-7(a))

$$\begin{aligned} \mu(POC_1)_{(low)} &= 0.8, \\ \mu(POC_1)_{(medium)} &= 0.2 \\ \mu(POC_1)_{(high)} &= 0.0 \text{ (not relevant to show in figure).} \end{aligned}$$

Where μ is the fuzzy membership function.

Similarly the membership values for other four inputs are :

$$POC_2 = 41 \text{ having } \mu(POC_2)_{(low)} = 0.0, \mu(POC_2)_{(medium)} = 1.0 \text{ and } \mu(POC_2)_{(high)} = 0.0$$

$$POC_3 = 23 \text{ having } \mu(POC_3)_{(low)} = 0.48, \mu(POC_3)_{(medium)} = 0.52 \text{ and } \mu(POC_3)_{(high)} = 0.0$$

$$POC_4 = 13 \text{ having } \mu(POC_4)_{(low)} = 1.0, \mu(POC_4)_{(medium)} = 0.0 \text{ and } \mu(POC_4)_{(high)} = 0.0$$

$$POC_5 = 9 \text{ having } \mu(POC_5)_{(low)} = 1.0, \mu(POC_5)_{(medium)} = 0.0 \text{ and } \mu(POC_5)_{(high)} = 0.0$$

The membership values for other four inputs are shown in figure 7(b-e).

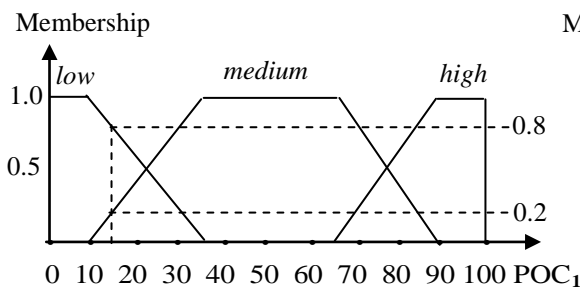


Fig 7 (a). Membership values of POC₁

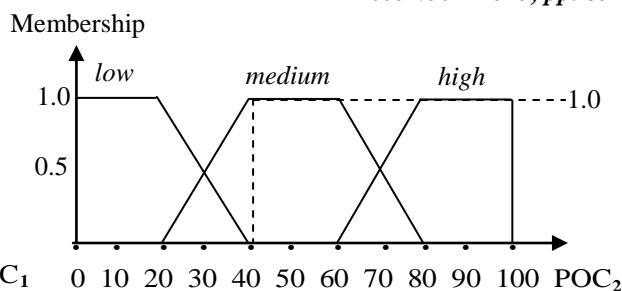


Fig 7(b). Membership values of POC₂

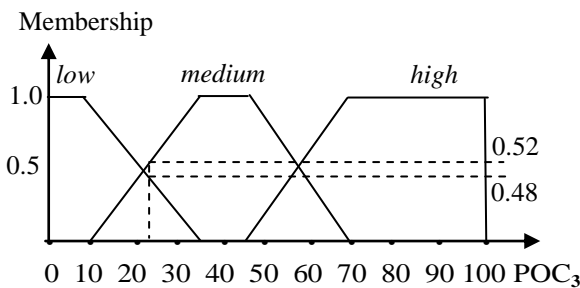


Fig 7(c). Membership values of POC₃

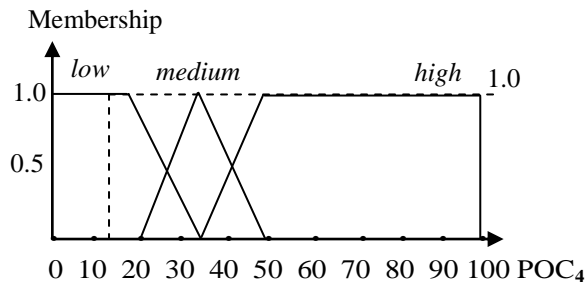


Fig 7(d). Membership values of POC₄

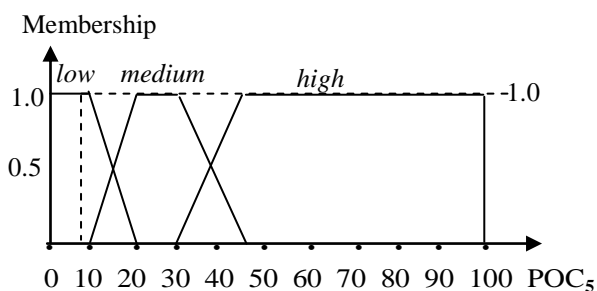


Fig 7(e). Membership values of POC₅

These input dataset activate four rules from rule base:

Rule 1. IF POC₁ is *low* AND POC₂ is *medium* AND POC₃ is *low* AND POC₄ is *low* AND POC₅ is *low* THEN EI is *low*.

The output (EI) is then set to *low* but to a degree of membership given by

$$\begin{aligned} \mu(EI)_{low} &= \text{Min}(\mu(\text{POC}_1)_{low}, \mu(\text{POC}_2)_{medium}, \mu(\text{POC}_3)_{low}, \mu(\text{POC}_4)_{low}, \mu(\text{POC}_5)_{low}) \\ &= \min(0.8, 1.0, 0.48, 1.0, 1.0) \\ &= 0.48 \end{aligned}$$

Rule 2. IF POC₁ is *medium* AND POC₂ is *medium* AND POC₃ is *low* AND POC₄ is *low* AND POC₅ is *low* THEN EI is *low*.

The output (EI) is then set to *low* but to a degree of membership given by

$$\begin{aligned} \mu(EI)_{low} &= \text{Min}(\mu(\text{POC}_1)_{medium}, \mu(\text{POC}_2)_{medium}, \mu(\text{POC}_3)_{low}, \mu(\text{POC}_4)_{low}, \mu(\text{POC}_5)_{low}) \\ &= \min(0.2, 1.0, 0.48, 0.0, 0.0) \\ &= 0.0. \end{aligned}$$

Rule 3. IF POC₁ is *low* AND POC₂ is *medium* AND POC₃ is *medium* AND POC₄ is *low* AND POC₅ is *low* THEN EI is *medium*.

The output (EI) is then set to *medium* but to a degree of membership given by

$$\begin{aligned} \mu(EI)_{medium} &= \text{Min}(\mu(\text{POC}_1)_{low}, \mu(\text{POC}_2)_{medium}, \mu(\text{POC}_3)_{medium}, \mu(\text{POC}_4)_{low}, \mu(\text{POC}_5)_{low}) \\ &= \min(0.8, 1.0, 0.52, 1.0, 1.0) \\ &= 0.52. \end{aligned}$$

Rule 4. IF POC₁ is *medium* AND POC₂ is *medium* AND POC₃ is *low* AND POC₄ is *low* AND POC₅ is *low* THEN EI is *low*.

The output (EI) is then set to *low* but to a degree of membership given by

$$\begin{aligned} \mu(EI)_{medium} &= \text{Min}(\mu(\text{POC}_1)_{medium}, \mu(\text{POC}_2)_{medium}, \mu(\text{POC}_3)_{low}, \mu(\text{POC}_4)_{low}, \mu(\text{POC}_5)_{low}) \\ &= \min(0.2, 1.0, 0.48, 1.0, 1.0) \\ &= 0.2. \end{aligned}$$

The numeric values corresponding to the output fuzzy labels are calculated at the maximum ‘truths’ of various rule strengths fed in from the rule evaluation step. The membership value for EI of fuzzy output ‘low’ = Max (0.48, 0.0, 0.2) = 0.48 and that is for EI of fuzzy output ‘medium’ = Max(0.52) = 0.52. The final output is obtained by combining the output as shown in figure 8.

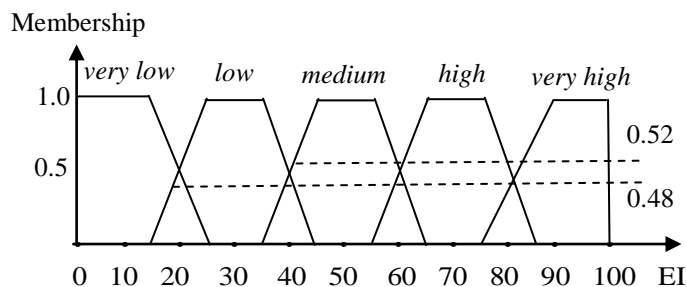


Fig 8. Fuzzification of the output variable EI

Based on the judgment of the experts, rules may be pruned out or added for more accuracy. For example, for rural area, where POC_1 is high, then all other four variables POC_2 , POC_3 , POC_4 and POC_5 can not be high simultaneously because it is contradictory in a real society. So this rule may be pruned by the experts.

The above four rules describe the output (EI) for the observed values of the system state variables. The final result is obtained by combining the outputs of the four rules and computing the fuzzy centroid.

When each rule is evaluated, the minimum numeric value of its antecedents is taken to be the rule strength. However, when determining the output fuzzy labels, the maximum numeric value is taken to be the labels value (“Min-Max” inference method).

Now defuzzification can be performed on these two labels (‘low’ and ‘medium’) to determine the actual real-world output (fuzzy centroidal value) of Education Index in terms of crisp value which is equal to 43.2. That is for a given Education Profile $\epsilon_p = \{14, 41, 23, 13, 9\}$, the value of Education Index is 43.2.

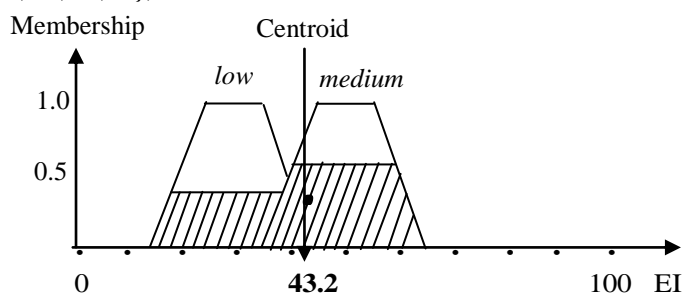


Fig 9. Defuzzification of the output variable EI

V. CONCLUSION

The level and spread of education has been an important precondition for sustained economic growth and it has also played a critical facilitative role in the demographic, social and political transition of these societies. It is a basic component of human development. So, fixing an index and measuring the value of educational attainment is an important issue. But, to the best of my knowledge, no work has been reported to properly address the issue. As fuzzy logic offers a natural and convenient framework in modeling various concepts in social science domain, so utilization of fuzzy logic is a graceful solution in measuring educational attainment.

The model has been developed by using Fuzzy Inference System of MatLab 5.0 and contains 147 rules. To test the efficiency of the model, the Education Indices have been calculated with Education Profiles obtained from randomly chosen twenty five areas (villages and municipal wards). The result is as per expectation, the value of Education Index raises with better economic status (in terms of employment, land value, household goods, live stocks etc.) and vice versa. Due to the lack of space, the details data are not furnished here. It is an initial effort, the system requires better tuning.

REFERENCES

- [1] National Human Development Report, 2001, Planning Commission, Govt. of India, March, pp. 49-50. 2002.
- [2] Koushik Basu and James E. Foster. “On Measuring Literacy”. *Economic Journal*, Vol. 108, November, 1998.
- [3] L. Dutta. “Generalized Measure of Literacy”. mimeo, University of East Anglia, Norwich. 2002.
- [4] S. E. Green, T. A. Rich and E. G. Nesman. “Beyond Individual Literacy: The Role of Shared Literacy for Innovation in Guatemala”. *Human Organization*, Vol. 44, pp. 313-321. 1985.
- [5] P. M. Valenti. “Should We be Concerned about the Distribution of Literacy Across Households ? An Approximatic Investigation”. mimeo, Cornell University. 2001.
- [6] K. Basu, A. Narayan and M. Rovallion. “Is Literacy Shared Within Households ? Theory and Evidence from Bangladesh”. *Labour Economics*, Vol. 8, pp. 649-65. 2002.

- [7] K. Basu, J. E. Foster and S. Subramanian. "Isolated and Proximate Illiteracy and Why These Concepts Matter in Measuring Literacy and Designing Education Programme". *Economic and Political Weekly*, 35, 1 and 2, pp. 35-39. 1999.
- [8] S. R. Chakraborty and A. Majumder. "Issues in the Measurement of Literacy". mimeo, Indian Statistical Institute, Kolkata, 2001.
- [9] S. Mishra. "Isolated and Proximate Illiteracy". *Economic and Political Weekly*, 36, 22, pp. 2003-8. 2001.
- [10] S. Subramanian. "Equity, Efficiency and the Measurement of Literacy; Some Extension of the Basu-Foster Framework". mimeo, Madras Institute of Development Studies. 2001.
- [11] G. Psacharopoulos and A. M. Ariagada. "The Educational Composition of the Labor Force: an International Comparison". *International Labor Review*, 125, pp. 561-574, 1986.
- [12] L. Lau, D. Jamison and F. Louat. Education and Productivity in Developing Countries: an Aggregate Production Function Approach. Report No. WPS 612, The World Bank. 1991;
- [13] R. J. Barro and J. W. Lee. "International Comparisons of Educational Attainment". *Journal of Monetary Economics*, 32, pp. 363-394, 1993;
- [14] V. Nehru, E. Swanson and A. Dubey. "A New Data Base on Human Capital Stock: Sources, Methodology, and Results". *Journal of Development Economics*, 46, pp. 379-401, 1995.
- [15] M. Braun and W. Muller. "Measurement of Education in Comparative Research". *Comparative Social Research*, 16, pp. 163-201. 1997.
- [16] A. C. Kerckhoff and M. Dylan. "Problems with International Measures of Education". *Journal of Socio-Economics*, 28 (6), pp. 759-775. 1999.
- [17] W. Muller, P. Luttinger, W. Konig and W. Karle. Class and Education in Industrial Nations. In M. Haller (Ed.) Class structure in Europe. New Findings from East-West Comparisons of Social Structure and Mobility, Armonk (NY), London: M.E. Sharpe. 1990.
- [18] UNESCO : Operational Manual for ISCED 1997 (International Standard Classification of Education). Paris: UNESCO, First edn. 1999.
- [19] UNESCO Institute for Statistics (2006 [1997]). International standard classification of education: ISCED 1997 (re-edition). Montreal: UNESCO Institute for Statistic, URL http://www.uis.unesco.org/TEMPLATE/pdf/isced/ISCED_A.pdf.
- [20] S. L. Schneider. "Measuring Educational Attainment in Cross-National Surveys: The Case of the European Social Survey". Proc. of 8-th Conf. of the European Sociological Association (ESA), Glasgow, 3-6 September, 2007.
- [21] L. A. Zadeh. "Fuzzy sets and systems." In: Fox, J. Ed., *System Theory*. Polytechnic Press, Brooklyn, NY. pp. 29-37. 1965.
- [22] Charles C. Ragin. *Fuzzy-Set in Social Science*. Chicago: The University of Chicago Press. pp. 4-5. 2000.
- [23] C. Ponsard. "An application of fuzzy subsets theory to the analysis of consumer's spatial preference." *Fuzzy Sets and Systems*, 5(3), pp. 235-244. 1981.
- [24] C. Ponsard. "Fuzzy mathematical models in economics." *Fuzzy Sets and Systems*, 28(3), 1988. pp.273-283.
- [25] *Fuzzy Sets and Systems*. Vol. 49, No. 1, 1992.
- [26] G. Q. Chen, S. C. Lee and E. S. H. Yu. "Application of fuzzy set theory to economics." In: Wang, P. P., Ed., *Advances in Fuzzy Sets, Possibility Theory, and Applications*. Plenum Press. New Work, pp. 277-305. 1983.
- [27] A. Billot. *Economic Theory of Fuzzy Equilibria: An Axiomatic Analysis*. Springer-Verlag, New York. 1992.
- [28] A. Cerioli and S. Zani. "A Fuzzy Approach to the Measurement of Poverty". In: Dagum, C. and Zenga, M. (eds), *Income and Wealth Distribution, Inequality and Poverty*. Berlin: Springer Verlag. 1990.
- [29] E. C. Martinetti. "Capability Approach and Fuzzy Set Theory: Description, Aggregation and Inference Issues". In: Lemmi, A. and Belli, G. (eds), *Fuzzy Set Approach to Multidimensional Poverty Measurement*. New York, Springer Verlag. 2006.
- [30] E. Cox. "Fuzzy Fundamentals". *IEEE Spectrum*, October, pp. 58-61. 1992.