

## Fuzzy Decision Tree System for Potential Evaluation of Land Consolidation

Jinfeng Wang<sup>1a</sup>, Aixia Ge<sup>2</sup>, Cheng Li<sup>2</sup>, Wenzhong Wang<sup>3</sup>, Yueming Hu<sup>2</sup>

<sup>1</sup>Department of Mathematics and Informatics, South China Agricultural University, Guangzhou, 510642, China

<sup>2</sup>Key Laboratory of the Ministry of Land and Resources for Construction Land Transformation, Guangzhou, 510642, China

<sup>3</sup>Department of economics and management, South China Agricultural University, Guangzhou, 510642, China

<sup>a</sup>wangphoenix@163.com

**Keywords:** Fuzzy Decision Tree, land consolidation, potential evaluation

**Abstract.** The traditional evaluation methods for land consolidation potential mainly depend on the experts' experience, statistics computations or subjective adjustments. There usually exists some bias in the results. So, the computer technology has been essential. In this study, an intelligent evaluation system based on Fuzzy Decision Tree is established, which can be deal with numerical data, discrete data and symbol data. When the land original data is input, its potential to be developed will be output by this new model. It is more helpful for authority to give out the objective proof for decision making. The project of Shunde's land consolidation provides the data support. The conventional evaluation results are compared. They are consistent roughly. The new model is more automatic and intelligent.

### Introduction

Rural conditions have been destroyed in many countries of the world and will maybe continue to worsen. Land consolidation(LC) can be an effective instrument in rural development. Land comprises two main components: land reallocation and agrarian spatial planning. Land reallocation can be called as land readjustment, which involves the rearrangement of ownership in terms of parcels (size, shape and location) and rights (land exchange). It is the core part of the land consolidation approach. Agrarian spatial planning includes the provision of the necessary infrastructure such as roads, irrigation systems, drainage systems, landscaping, environmental management, and village renewal and soil conservation [1]. LC aims to increase land process inefficiency [2, 3] and support rural development [4]. So, LC is very important for rural developing. How to precede the land consolidation and how to evaluate the land potential for consolidation are the crucial problems for authorities.

As so far, many researchers have been focus on the potential evaluation in the world. The Turkish statistical Institute (TUIK) performed a general agricultural census in Turkey decennially[5]. LC Projects were developed depending on experience[6,7,8]. The intelligent systems can interpret the professional result and enhance the cognitive performance of decision makers. A fuzzy expert system was proposed for analyzing and solving uncertainty in land data[9]. A Spatial Decision Support System (SDSS)-based land reallocation model was developed to provide reallocating newly created regular size parcels to landowners in land consolidation projects[9]. The models are constructed depending on computer technology, which are faster and more trustworthy. The results are still not intuitive and natural. In this paper, a fuzzy decision tree system for LC is proposed. The characteristics of Decision tree include strong interpretability, high accuracy and rapid implementation, which are compensate for the traditional models.

This article is organized as follows. The land consolation and some traditional methods have been introduced in this section. Section 2 contains the preliminaries, definitions, and terminologies needed for later sections. The next section gives the details of data from Shunde's project. Then

experiments provide the results and the analysis is proceeded in section 4. Summary and conclusions are provided in section 5.

## Method and model

The fuzzy logic was proposed in 1965 by Zadeh[10], which can describe and handle the vague and ambiguous data. Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values) fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Furthermore, when linguistic variables are used, these degrees may be managed by specific functions. Irrationality can be described in terms of what is known as the fuzzjective. Fuzzy logic has been applied to many fields, from control theory to artificial intelligence.

**Fuzzy set theory.** Fuzzy set theory is primarily concerned with quantifying and reasoning using natural language in which words can have ambiguous meanings. This can be thought of as an extension of traditional crisp sets, in which each element must either be in or not in a set. Fuzzy sets are defined on a non-fuzzy universe of discourse, which is an ordinary sets [11]. A fuzzy set is characterized by a membership function  $\mu_F(x)$  which assigns to every element, a membership degree  $\mu_F(x) \in [0,1]$ . When  $\mu_A(x) > 0$ , an element  $x \in U$  will be in a fuzzy sets  $F$ . especially,  $\mu_F(x) = 1$  represents a full member [12]. Membership functions can either be chosen based on the user's experience or by using optimization procedures[13][14]. Typically, a fuzzy subset  $A$  can be

represented as  $A = \left\{ \frac{\mu_A(x_1)}{x_1}, \frac{\mu_A(x_2)}{x_2}, \dots, \frac{\mu_A(x_n)}{x_n} \right\}$

Fuzzification is the process of changing a real scalar value into a fuzzy value [15]. This is achieved with the different types of fuzzifiers. In this paper, we adopted the Trapezoidal or triangular fuzzifier. Fuzzification of a real-valued variable is done with intuition, experience and analysis of the set of rules and conditions associated with the input data variables. There is no fixed set of procedures for the fuzzification.

**Fuzzy Decision Tree Construction.** Fuzzy sets and fuzzy logic are able to deal with the language-related uncertainties by fuzzifying, while providing a symbolic framework for increasing knowledge comprehensibility. Fuzzy decision trees differ from traditional crisp decision trees in three respects [16]: splitting criteria based on fuzzy restrictions, the different inferring procedures and defining the fuzzy sets representing the data. The heuristic for the fuzzy decision tree is based on the minimal ambiguity.

The procedure of constructing FDT is mainly as follows.

Place all data into one node as root;

Select one feature with low entropy to divide the cases in the root into different son-nodes according to the different feature values;

For each son-node, repeat the same action until the node cannot be divided, i.e. leaf.

Given that non-leaf nodes  $S$  has  $n$  fuzzy features  $A^{(1)}, A^{(2)}, \dots, A^{(n)}$  to be selected, for every  $k$  ( $1 \leq k \leq n$ ), fuzzy feature  $A^{(k)}$  takes  $m_k$  linguistic values as  $T_1^{(k)}, T_2^{(k)}, \dots, T_{m_k}^{(k)}$ .  $A^{(n+1)}$  represents class which takes values as  $T_1^{(n+1)}, T_2^{(n+1)}, \dots, T_m^{(n+1)}$ . In symbolic datasets, the value of features and classes are 0 or 1. For better description, we define  $|S|$  represents the number of examples of the non-leaf node  $S$ .

For every value of attribute,  $T_i^{(k)}$  ( $1 \leq k \leq n, 1 \leq i \leq m_k$ ), the relative frequency about the  $j^{\text{th}}$  class  $T_i^{(n+1)}$  on non-leaf node  $S$  is defined as  $p_{ij}^{(k)} = |S_i \cap S_j| / |S_i|$ , in which  $S_i$  is the subset of  $S$  for

which attribute  $A^{(k)}$  has value  $T_i^{(k)}$  (i.e.,  $S_i = \{s \in S \mid A^{(k)} = T_i^{(k)}\}$ ) and  $S_j$  is the subset of  $S$  too for which  $A^{(n+1)}$  take value  $T_j^{(n+1)}$  (i.e.,  $S_j = \{s \in S \mid A^{(n+1)} = T_j^{(n+1)}\}$ ). On non-leaf node  $S$ , the classification entropy of  $T_i^{(k)}$  is defined as  $Entr_i^{(k)} = -\sum_{j=1}^m |S_i \cap S_j| / |S_i| \log_2 |S_i \cap S_j| / |S_i|$ .

The average classification entropy of the  $k$ th attribute is defined as  $E_k = \sum_{i=1}^{m_k} \omega_i Entr_i^{(k)}$ , in which  $\omega_i$  represents the weight of the  $i^{\text{th}}$  value  $T_i^{(k)}$ ,  $\omega_i = |S_i| / |S|$ . So we can summarize to get the entropy, i.e.,  $E_k = \sum_{i=1}^{m_k} \frac{|S_i|}{|S|} Entr_i^{(k)}$ .

FDT aims to find out one attribute which can make the average classification entropy minimum, i.e., selecting one integer  $k_0$  so that  $E_{k_0} = \text{Min}_{1 \leq k \leq n} E_k$ .

## Data preparation

In this study, we took the ‘three old’ reconstruction project of City Shunde of Province Guangdong in China as a research case. The ‘three old’ means old villages, old factories and old towns. Shunde is the pioneer in economic reformation in Guangdong. The government proposed the ‘there old’ consolidation project to strengthen the management of constructive land and promote insulating the intensive land.

**Data description.** The potential evaluation of ‘three old’ land consolation mainly focus on those land blocks which have contained into the plotting and building database. The statistical data is shown as table 1. This project is characteristic as big area, quantity, wide range and concentrated distribution.

**Table 1 Statistical data of the types and area of ‘three old’ reconstruction**

Type state	Old factory	Old town	Old villages	In total
finished	4437.91	568.27	44.68	5050.86
On-going	4196.71	354.35	563.51	5114.57
Not starting	49634.37	9074.25	8425.73	67134.35
In total	58268.99	9996.86	9033.92	77299.77

**Data pre-process.** In this project, the evaluation targets are characteristic of multiple features. It is necessary to normalize all feature values for cancelling the influence from variables and values. One general method is 0-1Normalization to scaling the feature by bring all values into the range[0,1]. It is also called unity-based normalization.

Let  $X_{max_{ij}}$  indicts the maximum value and  $X_{min_{ij}}$  indicts the minimum value for the  $j^{\text{th}}$  feature of the  $i^{\text{th}}$  case. The normalization for each variable can be computed as follows.

$$\text{For active index: } S_{ij} = \frac{X_{ij} - \min X_{ij}}{\max X_{ij} - \min X_{ij}}; \text{For negative index: } S_{ij} = \frac{\max X_{ij} - X_{ij}}{\max X_{ij} - \min X_{ij}}$$

According to previous formula, the range of  $X'$  is between 0 and 1. The distribution of Each  $X'$  is same to the original value of  $X$ . The advantage of 0-1 normalization is that the best is always 1 and the worst is 0 whatever the value is negative or active. But it neglects the difference of the features’ value, which lead to missing the relationship among features. However, the 0-1 normalization is still the simplest method.

## Model and analysis

Before building the evaluation model, we need finish the feature selection to reduce the complexity of computation by deleting the redundant information. We adopt the WEKA exploit platform to call feature selection function and develop the evaluation model. After completing feature selection,

fuzzy decision tree is constructed for the preprocessed data for evaluating the comprehensive potential. The data of Shunde's project contains 477 blocks in which 27 blocks have been finished reformation and can be used as training set.

The model construction can be presented as figure 1.

In this project, those blocks which have been finished and are on-going transformation present their actual potential and are used as training set. The remainder which contains un-started ones are tested to compare with the conclusions which have drawn by statistic and analysis. All artificial marks are removed off form original data. The final dataset contains 27 predictive features and 3 levels of potential. Level one means the highest potential; level 2 represents the medium type; and level 3 is the worst grade for transformation. All results are listed in table 2 to show the situation of predicted potential of each town.

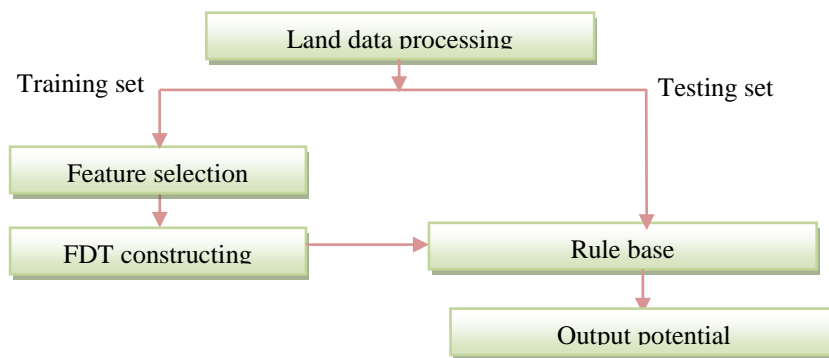


Figure 1 Flowchart of model construction

Assumed that the potential marked by experience is the destination classification, the prediction results of Fuzzy Decision Tree which is 89.12% shows high consistency with the artificial remarks and the actual land situation of Shunde's 'Three old' project. There is no block with level 1. It illustrates that there are not very old and battered buildings in Shunde district. In all blocks there exists level 2 and level 3. Those blocks in second ranking are characteristic as effective land use rate and modest volume rate. But due to bad living environment and ordinary location condition the price will not increase greatly. The third level blocks present reasonable volume rate and buildings density and good environment quality. Some basic facilities need to be improved, so the transformation potential is not so high. Longjiang, Lecong and Ronggui are arranged at the top three towns according to the ratio of level 2 which need to be transformed as key targets.

Table 2 potential level of each town

District	Number of Blocks	Level 1	Level 2	Level 3	Ratio of Level 2
Daliang	55	0	8	47	14.55
London	55	0	7	48	12.73
Ronggui	59	0	13	46	<b>22</b>
Leliu	44	0	7	37	15.9
Lecong	62	0	14	48	<b>22.58</b>
Junan	11	0	2	9	18.18
Longjiang	99	0	27	72	<b>27.27</b>
Beijiao	25	0	3	22	12
Chencun	15	0	3	12	20
Xingtian	26	0	2	24	7.69

## Conclusion

As so far, 'three old' transformation project is just developed in Guangdong, China. Research on 'three old' project is useful for international consolidation field. But the related study about

potential of transformation is very few. The traditional potential evaluation mostly depended on statistical method and experts' experiences. In this paper, a classical soft computing method---Fuzzy Decision Tree is induced to evaluate the potential of blocks for transformation. The results are more scientific, explicable and intelligent. The assessment of potential by FDT highly approves the conclusions drawn in traditional way. This study can provide an accessory support for decision making.

In addition, 'three old' transformation is a kind of policy problem which is affected by human factors. We need find more optimal methods to avoid of subjectivity. Meanwhile, there are too many features for one case of land project. Some is noisy information which is not good for model construction and final assessment. So indexes screening or feature selection is essential part of land consolidation.

## Acknowledgements

This research is supported by Ministry of Key Projects in the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period(No. 2013BAJ13B05), the National Natural Science Foundation of China(No. 61202295), and the National Social Science Foundation of China (Projects No. 10CJY024).

## Reference

- [1] J. Thomas. What's on regarding land consolidation in Europe? In: Proceedings of the XXIII International FIG Congress, Munich, Germany, October 8–13(2006).
- [2] P.M. Blaikie, A.Z. Sadeque: Policy in the High Himalayas: Environment and Development in the Himalayan Region. ICIMOD, Kathmandu(2000).
- [3] G.S. Niroula, G.B. Thapa: Impact of land fragmentation on input use, crop yield and production efficiency in the mountains of Nepal. *Land Degradation and Development* 18, 237–248(2007).
- [4] P. Sklenicka: Applying evaluation criteria for the land consolidation effect to three contrasting study areas in the Czech Republic. *Land Use Policy* 23 (4), 502-510(2006).
- [5] TUIK. Tarm Saym Sonuclar (in Turkish) <tuik.gov.tr>. (2001)
- [6] J. Sonnenberg: The European Dimensions and Land Management-Policy Issues (Land readjustment and land consolidation as tools for development), FIG Commission 7, 1996, Hungary.
- [7] J. Thomas: Actual trends concerning land management, land readjustment and land consolidation in Europe. In: Report at the 7th Workshop and 8th MC Meeting of the Action G9 of COST, Greece(2005).
- [8] J. Thomas. Property rights, land fragmentation and the emerging structure of agriculture in Central and Eastern European countries. *Electronic Journal of Agricultural and Development Economics Food and Agriculture Organization* 3, 225–275(2006).
- [9] Tayfun Cay, Fatih Iscan, Fuzzy expert system for land reallocation in land consolidation, *Expert Systems with Applications*. Volume 38, Issue 9, Pages 11055–11071, September(2011).
- [10] L. Zadeh. Fuzzy sets. *Information and Control*, 8, 338–353(1965).
- [11] Tien-Chin Wang and Hsien-Da Lee, Constructing a Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy, *WSEAS Transactions on Information Science and Applications*, Issue 8, Volume 3, pp.1547-1552, August(2006).
- [12] H. J. Zimmermann, *Fuzzy Set Theory and Its Applications*, Kluwer Academic Publishers, (1991).

- [13]J. S. R. Jang. Self-Learning Fuzzy Controllers Based on Temporal Back-Propagation, IEEE Trans. On Neural Network, Vol. 3, September, 714-723(1992).
- [14]S. Horikowa, T. Furahashi and Y. Uchikawa, On Fuzzy Modeling Using Fuzzy Neural Networks with Back-Propagation Algorithm, IEEE Trans. on Neural Networks, Vol.3, pp. 801-806, Sept.(1992).
- [15]L. H. Tsoukalas, R. E. Uhrig, Fuzzy and Neural Approaches in Engineering, John Wiley & Sons, Inc. (1993).
- [16]C. Z. Janikow, Fuzzy Decision Trees: Issues and Methods, IEEE Trans. on Systems, Man, and Cybernetics -Part B, Vo1.28, No.1, pp.1-14(1998).