

Feature Selection Based on Function Model of Game Fusion

Lin Gu

School of Computer Engineering and Science
Shanghai University
Shanghai 200444, Peoples Republic of China
gulin525@shu.edu.cn

Shardrom Johnson

Information Centre
Shanghai Municipal Education Commission
Shanghai 200003, Peoples Republic of China
jshardrom@shmec.gov.cn

Yang Zeng

School of Computer Engineering and Science
Shanghai University
Shanghai 200444, Peoples Republic of China
Zengyang123@shu.edu.cn

Daniel Hsu

School of Computer Engineering and Science
Shanghai University
Shanghai 200444, Peoples Republic of China
zoen@shu.edu.cn

Abstract—The paper introduces a new method of feature selection through the function model of game fusion. We use the traditional three steps of accomplishing feature selection by the fusion model: Firstly, fusion model can preprocess the input information, classify the information into conflict data, redundant data and complementary data, and filter the conflict data. Then, it transmits the preprocessed data to the modules of the information center. Finally, fusion model achieves the optimal situation through the fusion algorithm. We discover the complexity of these three steps will be high when the size of features is huge. Therefore, we draw on the experience of feature division and reduce the size of strategy set from power set of features to the features set.

Keywords—game theory; information fusion; feature selection

I. INTRODUCTION

Information fusion which is also named as data fusion is originated from sonar signal processing system of the United States Department of Defense in 1973. Information fusion merges the information from disparate sources which are always gathered by multi-sensor system.

The two critical elements of information fusion method are fusion model and fusion algorithm. There are plenty of methods accomplishing the information fusion in different field.

Information fusion based on game theory is a hot topic at present. Most of the methods based on game theory use the function model of game fusion which is proposed by Wang in [1], but the fusion algorithms are different. For example, Xiao takes the Bayesian network as the fusion algorithm [2]. Yin uses the Markov game theory and introduces a thought of using agent [3]. Xu uses Nash equilibrium which is a concept of economy [4]. Wei tries the technique of hotpot [5] and Guo introduces the way of game fusion using multi intelligent agents [6].

Feature selection is one of the applications can accomplish through information fusion based on game theory. We can achieve the goals of feature selection through

the game fusion model of [1] and the fusion algorithm of [4], but this attempt has a huge size of complexity. Fu introduces a new method of feature selection of game theory in [7]. Although the method of Fu has many limitations, we use the thought of this new method in game fusion model and achieve good effect finally.

II. INFORMATION FUSION BASED ON GAME THEORY

Information fusion process based on game theory is divided into three steps [9]:

Firstly, preprocessing gathered information to clean up incorrect information;

Secondly, input preprocessed data to the function model of the fusion system;

And then, the fusion algorithm is used in process refinement module of the fusion system, continuously optimize the strategy of each player, and ultimately form the optimal situation.

A. Data Preprocessing

To preprocess the input information, the concept of self entropy and mutual entropy which is mentioned in [8] should be used.

Giving n sensors, the measuring characteristics of each sensor S_i ($1 \leq i \leq n$) owns a probability distribution $p_i(x|x_i)$. Among them, $x \in X$ is random variables of measuring, x_i is the observed value of sensor S_i . $p_{ij}(x|x_i, x_j)$ is the combinational distribution of sensor S_i and S_j , x_i and x_j is the observed value of sensor S_i and S_j respectively [8].

The definition of self entropy is [8]:

$$h_i(x_i) = - \sum_{x \in X} p_i(x|x_i) \log p_i(x|x_i) \quad (1)$$

The definition of cross-entropy is [8]:

$$h_{ij}(x_i, x_j) = h_{ji}(x_j, x_i) = -\sum_{x \in X} P_{ij}(x | x_i, x_j) \log P_{ij}(x | x_i, x_j) \quad (2)$$

Data can be classified into redundant, complementary and conflict through the self entropy and mutual entropy. The conflict data is regarded as incorrect information must be errors occur in the sensors during collection. Thus, it should be filtered out before the data input into the fusion system.

The details of data classification as follows [8]:

There are six kinds of relations between $h_j(x_i, x_j)$ and $h_i(x_i)$, $h_j(x_j)$. Without loss of generality, three situations satisfy $h_i(x_i) < h_j(x_j)$ is considered:

1. $h_j(x_i, x_j) < h_i(x_i) \leq h_j(x_j)$.

In this situation, data are redundant, because the data of two sensors have positive correlation. For example, the sensors get the height and door height of a car will satisfy this situation, because door height strongly depends on the height of the car.

2. $h_i(x_i) \leq h_j(x_i, x_j) \leq h_j(x_j)$.

In this situation, data are complementary, because the data of two sensors have irrelevant or weak correlation. For example, the height and the weight of a student have weak correlation. Thus, the mutual entropy does not sacrifice a lot and keeps between two self entropies.

3. $h_i(x_i) \leq h_j(x_j) < h_j(x_i, x_j)$.

In this situation, data are conflicting, because the data of two sensors have negative correlation. The negative correlation implies that the sensor has got the incorrect data. Therefore, this situation can be applied as the process of preprocessing.

Summary of the relationship between entropy and data classification is shown in Table 1.

TABLE I. ENTROPY AND DATA CLASSIFICATION [8]

Relationship of mutual entropy and self entropy	Data correlation	Data classification
$h_j(x_i, x_j) < h_i(x_i) \leq h_j(x_j)$	Positive correlation	Redundancy
$h_i(x_i) \leq h_j(x_i, x_j) \leq h_j(x_j)$	Weak correlation	Complementary
$h_i(x_i) \leq h_j(x_j) < h_j(x_i, x_j)$	Negative correlation	Conflict

Through conflicting data filtering, the accuracy and reliability of fusion results can be improved.

B. Function Model of Fusion System

There are three crucial elements in the game theory[10]:

Player: Each side who participates in the decision is defined as a player. One decision can include two players, as well as more than two players. Players must have the strategy which is available to choose and can influence the final results.

Strategy: The feasible options which the player can take are defined as the strategy. The whole strategies of one player are a strategy set. Therefore, one strategy set corresponds to a player.

Define S_i as the strategy set of player i . When each side select a strategy from their own strategy set respectively, all these strategies can be noted as S , which is named as a situation.

Payoff: The vector function F which is based upon situation set S is defined as payoff function. For each situation S , $F(S)$ works out the payoff value through the strategy result of each player.

Therefore, a policy model is consisted of three parts: player, strategy and payoff.

The function model of fusion system includes[11]: Human-Computer Interaction System, Multi-Sensor System, Player Abstraction, Strategy Abstraction, Payoff Abstraction, Situation Assessment, Process Refinement, Database Management System. Figure1 shows the function model of game fusion.

Database Management System includes strategy database, payoff database. Process refinement module completes filtering conflict data and continuously choosing strategy to assess the situation. Situation assessment module calculates payoff value through the current situation.

The steps of function models are listed as follows[10]:

Step 1: Multi-Sensor System sends the collected data into information fusion center.

Step 2: Process refinement module filters the conflict data and takes the rest data as players.

Step 3: Strategy abstraction and payoff abstraction module get the set of strategies and the payoff functions from database management system through the current players.

Step 4: The process refinement module chooses one strategy from each player and forms a situation, then sends it to the situation assessment module.

Step 5: Situation assessment module calculates the payoff by situation and payoff functions and transmits the payoff to the process refinement module.

Step 6: Process refinement module judges whether the situation is the best. If not, go back to Step 4.

Step 7: Save the best situation in database management system and transmit it to human-computer interaction system.

C. Fusion Algorithm

Process refinement module needs to check whether the current situation is optimal. This step can reduce the computation of information fusion. There are many algorithms, such as Nash equilibrium [4], Bayesian network [2] and Markov game theory [3], etc.

Nash equilibrium is the most popular fusion algorithm owing to the concise and efficiency of it. It just judges the situation whether the players can increase their payoff just changing the strategy on his own[4]. In this paper, we take Nash equilibrium as the fusion algorithm of the process refinement.

III. FEATURE SELECTION BASED ON GAME THEORY

Using function model of game fusion can accomplish the application of feature selection. But when there are a great amount of features to select, the complexity of this function model confines it to achieve the goals. In [7], Fu has

proposed a method of feature selection based on game theory which can solve the problems of large size of features and has achieved a fantastic effect in vehicles recognition. In this paper, we analyze the method of [7], find the limitations of it and propose a new method which takes the advantages of these two methods.

A. Feature Selection Based on Game Fusion Model

Like the steps mention above, feature selection also can be done as follows[12]:

Step 1: Multi-sensor system transmits the collected data which contain the value of their features into the information fusion center.

Step 2: Filter conflict data and the rest data is the player.

Step 3: Set the power set of features as the strategy of each player.

For example, if the feature set of each player is {length, width, height}, the strategy is {{length}, {width}, {height}, {length, width}, {length, height}, {width, height}, {length, width, height}}.

Step 4: Each player chooses a same strategy which forms a situation.

Step 5: Assess the situation through summing up the different payoff of each player taking the strategy which is chosen.

Step 6: If the situation is Nash equilibrium, go to Step 7; if not, go to Step 4.

Step 7: Save and present the best situation.

Owing to the strategy is a power set of features, the complexity of this method is a problem.

B. Feature Selection Based on Game Theory and Feature Division

In [7], Fu has separated the features set into two parts and each part contains redundant features. Through constructing payoff function and choosing the payoff of the best situation, the author gets two features in each part[13].

In the experiment of vehicles recognition, the set of features is {length, width, wheelbase, front track, rear track}. It is obvious that length and wheelbase are a group of redundant features, while width, front track and rear track are another group of redundant features. The authors construct the player, strategy and payoff as follows[14]:

Set two players S_1 and S_2 .

Set the strategy of S_1 as {length, wheelbase}, S_2 as {width, front track, rear track}.

Set the payoff function [7]:

$$a_j = \frac{\sum_{k=1}^n p(\omega_k | \alpha_i) \log p(\omega_k | \alpha_i) + \sum_{k=1}^n p(\omega_k | \alpha_i) \log p(\omega_k | \alpha_i) - \sum_{k=1}^n p(\omega_k | \beta_j) \log p(\omega_k | \beta_j)}{3} \quad (3)$$

Where $\{\alpha_1, \alpha_2\} = \{\text{length, wheelbase}\}$, $\{\beta_1, \beta_2, \beta_3\} = \{\text{width, front track, rear track}\}$, α_j means the payoff of choosing the strategy (α_i, β_j) .

This method gives a new view of choosing players. It changes the strategy from the power set to the set of features, which greatly reduce the complexity. But it also has limitations which are listed as follows:

The method is based on a hypothesis that the features must be redundant. Otherwise, the features cannot be put in different sets.

The number of sets of redundant features must be two, which is unsuitable for many applications in real life. For instance, in vehicles recognition, there also can be a group of features like height and door height.

C. A New Method of Feature Selection Based on Function Model of Game Fusion

Referring to the method of [7], we change the player and traditional strategy mention before. After Step 2, we classify the features by putting the redundant features in a same set. Redundancy of features will improve the accuracy of feature selection, but won't be necessary. If the features are not redundant we can separate the features randomly and will also get a result of feature selection.

After classifying, we count the number of sets. The count is the players and the sets of features are strategies.

For instance, the set of whole features is $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \delta_1, \delta_2, \delta_3, \delta_4\}$, suppose that after classification which puts the redundant features in a same set, the sets are $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$, $\{\beta_1, \beta_2, \beta_3\}$, $\{\gamma_1, \gamma_2\}$, $\{\delta_1, \delta_2, \delta_3, \delta_4\}$. Then, they are the strategy sets of player S_1, S_2, S_3, S_4 .

In this method, we can also use the payoff functions of [7], which calculate the mutual information between two sets of redundant features. We sum up the mutual information of each two sets of a situation and use the function model to check the Nash equilibrium.

The advantage of this method is not only changing the power set of features to the set of features, which reduces the complexity, but also enabling using the fusion algorithm such as Nash equilibrium in function model to accelerate finding the best situation.

IV. CONCLUSION

In this paper, we introduce the method of feature selection which uses function model of game fusion and improve it by changing the mode of player and strategy. Finally, we find more advantages of this method than the method before.

Although we can select feature through the method of using function model, construct new payoff functions which are more suitable for feature selection and find the new fusion algorithm to accelerate finding the best situation are necessary in future working[15].

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grants No.61303097, and by the Graduate Innovation Fund Project of Shanghai University (SHUCX070037; SHUCX120105).

REFERENCES

- [1] Gang Wang, "Problems on the game fusion under the conflict environment of multiple source information", Ph. D dissertation, Northeastern University, P.R. China, 2004. (in Chinese)
- [2] Xinhua. Xiao, "Research on application of game theory in the information fusion", Proceedings of the 2nd International Conference on Computer Engineering and Applications, 2010, pp. 48-51.
- [3] Huilin Yin, Lei Wang, and Jing Nong, "Survey on game-theoretic information fusion", Proceedings of the 7th International Conference on Fuzzy Systems and Knowledge Discovery, 2010, pp. 2147-2151.
- [4] Nan Xu and Xianjia Wang, "An information fusion method based on game theory", Proceedings of the 2nd International Conference on Signal Processing Systems, 2010, pp. 195-198.
- [5] Lifeng Wei and Xiaobin Wang, "Research on Honeypot Information Fusion Based on Game Theory", Proceedings of the 2nd International Conference on Computer Research and Development, 2010, pp. 803-806.
- [6] Changgeng Guo, Luo Zhong, and Tao Liu, "A Study on Game Theory in Decision Interaction for Multi Intelligent Agents Based on Information Fusion", Proceedings of the International Conference on Computing, Information and Control, Berlin Heidelberg: Springer-Verlag, 2011, pp. 442-452.
- [7] Maixia Fu, Feiyu Lian, Qing Li, and Yuan Zhang, "A New Method of Feature Selection for Information Fusion Based on the Game Theory", Proceedings of the 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, China: WuHan, 2008, pp. 185-189.
- [8] Minhua Liu and Deyun Xiao, "Method of Multisensor Data Classification Based on Entropy", Control and Decision, 2006, vol. 21, no. 4, pp. 410-414, 420. (in Chinese)
- [9] Bienvenu Fassinut-Mombot and Jean-Bernard Choquel, "An Entropy Method for Multisource Data Fusion", Proceedings of the 3rd International Conference on Information Fusion, 2000, pp. 10-13.
- [10] Colin Farrell Camerer, "Progress in Behavioral Game Theory", Economic Perspectives, 1997, vol. 11, no. 4, pp. 167-188.
- [11] Wei Pan, Yangsheng Wang, and Hongji Yang, "New General Function Model of Multi-model Information Fusion System-Solution Based on Fusion Function and Information Hierarchy", Computer Engineering and Applications, 2006, vol. 42, no.29, pp. 27-29, 35 . (in Chinese)
- [12] Feiyu Lian, Maixia Fu, Yuan Zhang, and Ke Wang, "Investigation of Game-fusion Model for Multi-source Information Fusion and its Application in ETC", Journal of Lanzhou University of Technology, 2007, vol. 33, no.3, pp. 83-87. (in Chinese)
- [13] Xin Sun, Yanheng Liu, Jin Li, Jianqi Zhu, Huiling Chen, and Xuejie Liu, "Feature Evaluation and Selection with Cooperative Game Theory", Pattern Recognition, 2012, vol. 45, no. 8, pp. 2992-3002.
- [14] Shay Cohen, Gideon Dror, and Eytan Ruppim, "Feature Selection via Coalitional Game Theory", Neural Computation, 2007, vol.19, no. 7, pp. 1939-1961.
- [15] Shardrom Johnson, "Clustering Preprocessing of Data Resource", Shanghai Popular Science Press, 2011.

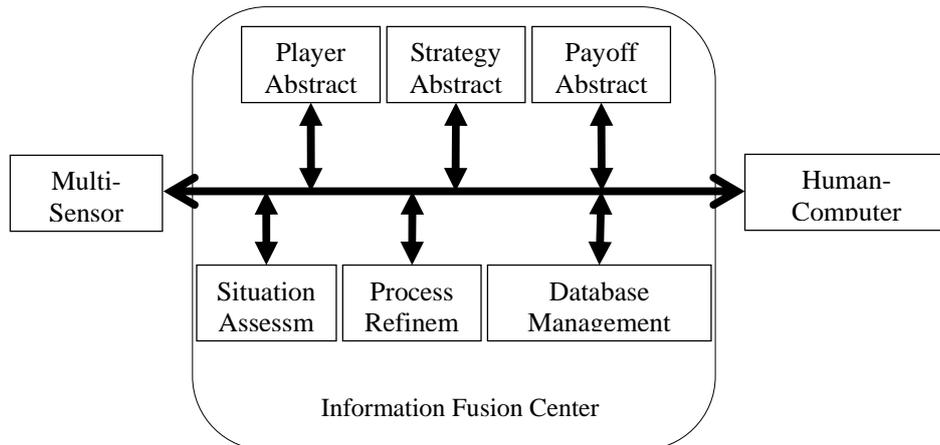


Figure 1. Function model of game fusion [1].