

# Mining Method of Recessive Lineage Relationship between Policies

Gang Liu, Hefei Wang<sup>\*</sup>, and Honglei Zhang

*School of Computer Science and Technology, Harbin Engineering University, Harbin, 15001, China*

---

## Abstract

Based on summarising and analyzing the research status of policy research and the technology of concept description, this paper proposes an effective mining method of recessive lineage relationship between policies. This method introduces the theory of factor space and gives a method for decomposing concept factor. We introduce the recessive gene as a new "synonym" and calculate the fitting degree of policy texts to reveal the recessive lineage relationship, which cannot be reflected through the direct calculation of similarity. Finally, we use the articles of law as the set of policy texts to carry out a large number of experiments. The comparison and analysis of the experimental results verify the effectiveness of the proposed method.

*Keywords:* potential similarity; factor space; extraction of commonalities; recessive lineage

(Submitted on April 3, 2018; Revised on May 1, 2018; Accepted on June 20, 2018)

© 2018 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

Inconsistencies in public policy can be found everywhere, and the social impact of these inconsistencies is enormous [2]. For example, the Law on Government Procurement stipulates that this law applies to institutions and state organs at all levels that use financial funds to purchase goods or services. These goods or services must be in accordance with the provisions of the procurement catalog and meet the quota standards [9,12,13]. At the same time, the Law on Tendering and Bidding applies to all tendering and bidding activities conducted in China. That means, regardless of the nature of the unit, the type of bidding objects and the source of funds, all related activities are regulated by the Law on Tendering and Bidding. As an integral part of the government procurement legal system, the two laws have the same legal effect, but there is a clear conflict in the sphere of the application of the two laws. Based on the conflict between these two laws, a series of behaviors have emerged that evade the other legal liability on the ground of only being bound by one of the laws. Such acts inevitably lead to chaos in the order of government procurement and reduce the effectiveness of the administrative system.

For the inspection of policy inconsistency, qualitative analysis is a major method and has achieved some results. However, given the numerous and constantly updated policies in various places, qualitative analysis can only solve a small number of problems. However, given that the number of policies is large and constantly updated, qualitative analysis can only solve a small number of problems. Purely artificial or semi-rational policy analysis is difficult to get accurate conclusions. The methods of traditional artificial policy analysis are not applicable. Faced with a huge amount of computation, we cannot completely solve the above problems without introducing a certain degree of automation technology.

In the field of policy research, policies have shown great complexity in the subject, environment and evolution. In response to the complexity of the policy, the theory of complex adaptive system has emerged. Among them, multi-agent technology implementation of a system accounts for a large proportion. For example, literature [4] studies a way of multi-agent interaction systems, and this is feasible to achieve multi-policies. Literature [10] built a dynamic inconsistency model to study. Through this study, we can get effective measures to solve the zero solution of inflation policy. Moreover, this measure has helped a lot in the formulation of economic policies.

---

<sup>\*</sup> Corresponding author.

E-mail address: 549619649@qq.com

Based on the theory of factor space, this paper proposes a method of mining the recessive lineage relationship between policies. Based on the concept of recessive lineage policy, this method applies the theory of factor space to the field of policy analysis for the first time and proposes a method of conceptual factor decomposition based on factor space. After that, we combine the results of conceptual factor decomposition with the calculation of the conventional discourse's similarity to reveal the implicit relationship among the policies by extracting the commonalities among the concepts. This method can help policy analysts effectively filter policy, provide some theoretical basis for policy analysts, and have an important reference for research in other related fields.

## 2. Related Work

### 2.1. Dominant and Latent Lineage Relation Between Policies

There are two objects to study on the issue of policy coherence, which are the internal consistency of policies and the consistency of different policies. The internal consistency of policies can be understood as the consistency between two identical policies. The internal consistency of policies is a special form of the consistency between policies, and the two are unified. Whether or not the policies are consistent can be traced back to the starting points of the policies. Because all laws and policies are derived from the constitution, all the existing laws have innate internal relations, that is, the lineage relationship of the policy [3,11].

**Definition 2.1.** The paternal kinship between policies described by the genealogy of the policy lineage relationship is denoted as Father Tie. Consistency and invariance of policy as blood ties continue to extend, diverge, and spread along tree structures. The relationship between the policies at the ends of the different branches is diminishing. However, we can occasionally find that there are strong similarities between some policies in different branches. This is rooted in the lineage relationship of the policies. In essence, it is the paternal kinship between policies. In the process of policy derivation, there is still the inheritance of the maternal and collateral kinship lineage, which cannot be represented by a lineage tree and became an implied kinship.

**Definition 2.2.** The kinship between policies that cannot be described by the genealogy of the policy lineage relationship is the maternal lineage relationship, which is denoted as Mother Tie. For the convenience of characterization, we refer to the kinship of the paternal lineage as the dominant lineage relationship and the kinship of the maternal lineage as the recessive lineage relationship. In fact, the inheritance and maintenance of kinship are achieved through familial genes [1]. The main transmission of policy kinship is policy genes.

**Definition 2.3.** The dominant gene of a policy refers to the policy gene inherited from the genealogy of its family, and the recessive gene of a policy refers to a new gene that is absorbed through maternal or collateral lineage during the process of policy derivation. Policies with a large number of identical recessive genes are likely to have very similar bloodlines in their maternal pedigree and have great research value in the study of policy consistency. Therefore, mining the recessive genes of policies has always been an important research point of policy consistency analysis.

### 2.2. Data Description

P.Z Wang proposed the concept of factor space [5] under the background that fuzzy set theory only provides the concept extension of knowledge representation but cannot describe the connotation of knowledge representation. Currently, the theory of factor space has been widely used in decision-making science, expert system, artificial intelligence, automatic generation of concepts, pattern recognition, and other related fields.

Factor space is based on the fact that anything in real life or production can be characterized by a combination of the factors involved, providing a mathematical description model for concepts [12]. By a combination of factors, things can be described using unique points in a multidimensional coordinate system. However, things do not relate to every single coordinate in a multidimensional coordinate system, leading to the following definition.

**Definition 2.4 [6].** Assuming that the domain (set of objects) is denoted as  $O$  and the set of factors is denoted as  $V$ , then for  $\forall f \in F$ , all the involved factors are in  $V$  for  $\forall o \in O$ , and all of the objects that are related to  $f$  are in  $O$ .  $(O, V)$  is called a pair.

**Definition 2.5 [5].** For a given pairing  $(O, V)$ , for any object  $o \in O$ ,  $R(o, f) = 1$  is true if and only if  $O$  relates to  $f$ , and then  $R$  is called the correlation.

**Definition 2.6 [6].** The factor  $f \in V$  is a mapping, for any  $o \in O$ , and the existing state  $f(o)$  corresponds to it:

- $f: D(f) \rightarrow X(f) \quad o \rightarrow f(o);$
- $X(f) = \{ f(o) \mid o \in O \}$

$X(f) = \{ f(o) \mid o \in O \}$  is called the state space of the factor  $f$ .

**Definition 2.7 [5, 15].** If the zero factor is the only true factor of factor  $f \in V$ , then  $f$  is called the atomic factor, and the set of all the atomic factors is called the atomic factor set, denoted as  $\pi$ , where each factor in  $\pi$  is mutually independent.

**Definition 2.8 [5, 6].** Factor space is a quaternion:

Discourse domain  $O$ ;

Set of state space  $X(f)$ ;

A mapping set from  $O$  to  $X(f)$ , which is  $f: O \rightarrow X(f), f \in V$  we call  $f \in V$  as a basic factor;

Let  $\mathfrak{R}$  denote the algebra  $\sigma$  formed by the subset of  $V$ , the element in  $\mathfrak{R}$  is  $A \subseteq V$ , then  $A$  is a composite factor; If

$A(o) = \prod_{f \in A} f(o)$  is a Cartesian product, then  $V$  is called the full factor; if  $1$  is the full factor, then  $X(1) = \prod_{f \in V} f(o)$

$X(1) = \prod f(o)$  is called full space and  $\Phi$  is zero factor.

**Definition 2.9 [7]:** For any factor  $f \in \mathfrak{R}$ , if  $o_1 \sim o_2 \Leftrightarrow f(o_1) = f(o_2)$  holds, the equivalence class of  $\sim$  is called the atomic concept determined by  $f$ , and the union of atomic concepts is called the composite concept and is denoted as  $D(f)$ .

According to the above definition, the concept representation based on factor space is derived. The definition of the description frame is: Suppose the set of the concept is  $e = \{\alpha, \beta, \gamma, A\}$ , its discourse domain is  $O$ , taking factor family  $V$ . Then  $O$  and  $V$  form the left pairing  $(O, V)$  and take the factor set  $F \subset V$ , so that  $F$  is sufficient for  $O$ , and  $(\forall o_1, o_2 \in O)(\exists f \in F). f(o_1) \neq f(o_2)$  then  $(O, e, F)$  is a description frame of the concept set.

The theory of factor space provides the mathematical basis for the description and characterization of policy genes.

### 3. The Mining Method of Recessive Lineage Relationship Among Policies

The minimum granularity of policy is made up of concepts. The mining method mainly utilizes the theory of factor space to factorize the concept in the factor space to realize the deeper characterization of the concept, thus mining the deep recessive lineage relationship between policies. Given the concepts in different policy texts, they can be divided into common concepts and unique concepts. Given the policy text  $A$  and policy text  $B$ , the concept set that consists of the concept contained in the policy text  $A$  and policy text  $B$  is  $ConS$ .

**Definition 3.1.** For any concept  $c \in ConS$ , if the policy text  $c \subset A$  and the policy text  $c \subset B$ , then the concept  $c$  is the common concept of policy  $A$  and policy  $B$ , which is referred to as the common concept.

**Definition 3.2.** For any concept  $c \in ConS$ , concept  $c$  is the unique concept of policy text  $A$  if policy text  $c \subset A$  and policy text  $c \not\subset B$ ; similarly, concept  $c$  is the unique concept of policy text  $B$  if policy text  $c \subset B$  and policy text  $c \not\subset A$ . This article refers to the unique concept of policy text  $A$  and the unique concept of policy text  $B$  as the unique concept.

#### 3.1. Characterization of Conceptual Factors

The "factor" is explained in the dictionary as "the element that constitutes the thing; the reason or the condition that determines the success or failure of the thing" [8]. Everything intersects the factors [14] so everything can be described by a vector in a generalized coordinate system. The key issue in establishing such a generalized coordinate system is to find out the common signs, that is, the factors. We call this coordinate system a factor space.

##### 3.1.1. Determine the basic factor set BFS

Based on the factors and the theory of factor space mentioned above, the process of selecting  $BFS$  is mainly based on the initial mergence in the domain and the categorization of concept meaning in the dictionary. This paper presents an algorithm to determine  $BFS$ .

**Algorithm 1** Algorithm to get *BFS***Input:** initialize  $BFS = \{\text{noun, verb, adjective, adverb}\}$ **Output:** get *BFS*

1. for  $c \in \text{ConS}$ ,
2. {for  $bf \in BFS$
3. { if  $c$  has a projection on  $bf_i$ , then  $c$  has a coordinate value of 1 on  $bf_i$
4. { else  $c$  has a coordinate value of 0 on  $bf_i$ ;
5. }
6. Extract a set of factors with a relatively high degree of distinction from  $c$ , and call the set as  $\text{newF}$ ;
7.  $BFS = BFS \cup \text{newF}$ ;
8. }
9. End

Figure 1. Algorithm for BFS

**3.1.2. Building the preferred factor space**

In *BFS* (see Figure 1), each factor has a different representational ability to represent the entire set of common concepts. If multiple concepts have a value of one on the dimension of the factor, it indicates that the factor has a strong ability to represent the entire set of common concepts. If only a few concepts have a value of one on the dimension of the factor, it indicates that the factor has a weaker ability to represent the common set of concepts. In this paper, *BFS* is divided into three sets, which are as follows:

- The preferred factor space PFS: The factors with higher statistical coverage play a decisive role in the characterization of the common concept sets, and the factor space formed by the factor sets mentioned above is the preferred factor space.
- The alternative factor space AFS: In the characterization of the common concept sets, the factors with medium statistical coverage have the function of complementing and strengthening the set of preferred factors, and the factor space formed by the factor sets mentioned above is the alternative factors space.
- The other factor space OFS: The factors with lower statistical coverage play a weak role in the characterization of the common concept sets, and the factor space formed by the factor sets mentioned above is the other factor space.

Factors in *AFS* and *OFS* are mainly used to expand *BFS*, and in the expansion, we give priority to factor  $f \in AFS$  and then consider  $f \in OFS$ . As shown in Figure 1. Because the research focus of this paper is not the complete description of the concept, we use the conceptual characterization of the factors to achieve the latter mining research of the recessive lineage. Therefore, without seriously impacting the characterization, this article will choose the preferred factor space. Pruning the factor space and introducing some factors as appropriate can greatly simplify the state space and greatly reduce the complexity of the algorithm.

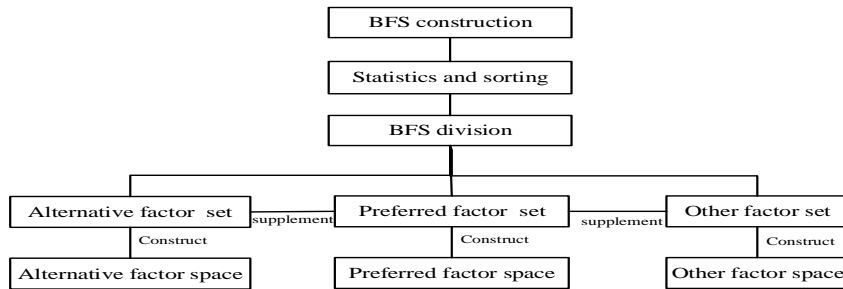


Figure 1. The basic process of constructing PFS

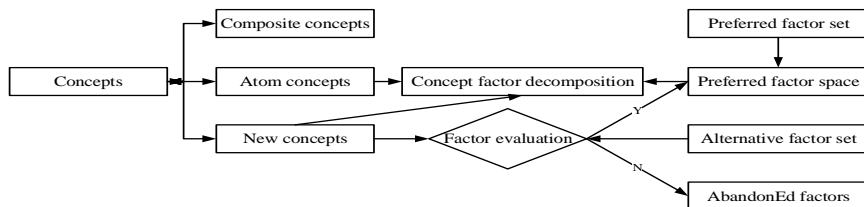


Figure 2. The basic process of conceptual factor's decomposition

**3.1.3. Decomposition of conceptual factors**

As shown in Figure 2, the decomposition of conceptual factors is a process of characterizing concepts in a factor space. The study suggests that in the process of characterizing factor, the concept sets *CS* can be divided into three subsets, which are

called atom concept set  $AC$ , compound concept set  $CC$  and new concept set  $NC$ .  $AC$ ,  $CC$  and  $NC$  do not intersect.  $V_c$  represents the conceptual decomposition vector of concept  $c$  and  $V_{ci}$  represents the component of the vector on the  $f_i \in BFS$ . Algorithm 2 can describe the basic process of conceptual factor decomposition.

---

**Algorithm 2** The decomposition of conceptual factor

---

**Input:** conceptual set  $CS$

**Output:** Conceptual factors after decomposition

1. Decomposing conceptual set  $CS = AC \cup CC \cup NC$ ;
2. For all  $f \in OFS$ ;
3. If  $f$  can describe most  $nc \in NC$
4.  $BFS = BFS \cup \{f\}$
5. For all  $c \in AC \cup NC$
6. For all  $f_i, i \in BFS$
7. If  $f$  can represent  $c$
8.  $V_{ci} = 1$
9. Else
10.  $V_{ci} = 0$
11. For all  $cc \in CC$
12. Decomposing  $cc$  to atom concept  $c_1, c_2, \dots, c_n$
13. For all  $ac_n$  do 5)~10)
14. For all  $V_{ci}$
15. If  $\sum_{i=1}^n V_{ci} \neq 0$
16.  $V_{cci} = 1$
17. Else
18.  $V_{cci} = 0$
19. For all  $f \in OFS$
20. If  $f$  can represent most  $nc \in NC$
21.  $BFS = BFS \cup \{f\}$
22. End

### 3.2. Extraction of Commonalities Between Policies

The commonalities between concepts are used to measure the strength of the relationship between concepts. Through the fitting operation of the conceptual factor vectors, we can find the connection between seemingly unrelated concepts. This deep level of connection is essentially the result of inheritance of recessive genes between concepts. But, not all of the common features extracted from the concept are recessive genes. Only those commonalities that confirm the existence of lineage relationships stabilize the final recessive gene.

**Definition 3.3.** In the factor space, the vector used to represent the commonality between concepts is called the commonality factor vector between concepts. The extraction of commonalities between policies is mainly composed of two steps: the calculation of vector matching and the calculation of commonalities, as shown in Figure 3.

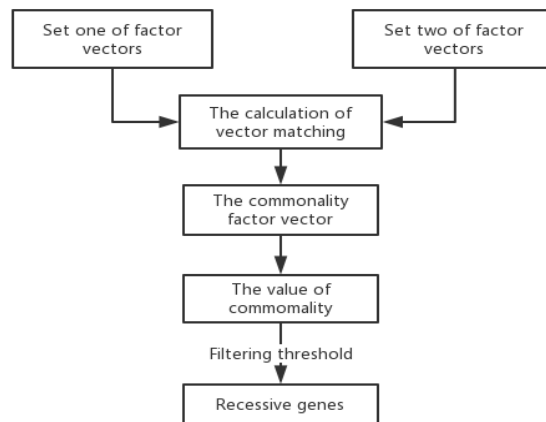


Figure 3. The process of extracting the commonalities between concepts

### 3.2.1. The calculation of vector matching

Suppose there is concept  $C_1$  and concept  $C_2$ . In the process of extracting the commonality between  $C_1$  and  $C_2$ , we need to consider their values on the same dimension  $f$ . The value of the common factor vector  $CFV$  in a certain dimension is determined jointly by the value  $cf_1$  of  $C_1$  in the dimension and the value  $cf_2$  of  $C_2$  in the dimension as follows:

- 1)  $cf_1 = cf_2 \neq 0$ , then the value of  $CFV$  in dimension  $f$  is 1;
- 2)  $cf_1 = cf_2 = 0$ , then the value of  $CFV$  in dimension  $f$  is  $\beta$  ( $0 < \beta < 1$ );
- 3)  $cf_1 \neq cf_2$ , and  $cf_1 \neq 0$ ,  $cf_2 \neq 0$ , but  $cf_1 \cap cf_2 \neq \emptyset$ , then the value of  $CFV$  in dimension  $f$  is  $\alpha$  ( $\alpha \in (0, 1)$ );
- 4)  $cf_1 \neq cf_2$ , and  $cf_1 \neq 0$ ,  $cf_2 \neq 0$ , but  $cf_1 \cap cf_2 = \emptyset$ , then the value of  $CFV$  in dimension  $f$  is 0.

### 3.2.2. The calculation of commonalities between policies

According to the theory of vector matching, we can get the value of  $CFV$  between each pair of concepts. The method of calculating the commonality based on  $CFV$  is as follows:

The commonalities between concepts are mainly expressed in the dimension of a non-zero value in  $CFV$ . The impact of dimension with a value of one,  $\beta$  or  $\alpha$  on the commonalities of concepts is not the same. When we calculate the commonalities, it should give different weights. Suppose the weight of a dimension with a value of one is set to be half the total number of dimensions, then the weight of a dimension with a value of  $\beta$  is one, and the value with the value of  $\alpha$  has a value that is between one and half the total number of dimensions. Therefore, in this method, the value of commonality is equal to the sum of all non-zero dimension weights in  $CFV$ , and the following equation is obtained:

$$x = c_1 * \frac{D}{2} + c_\beta + c_\alpha * \theta \quad (1)$$

Wherein,  $C_1$  represents the number of dimensions with a value of 1 in  $CFV$ ,  $c_\alpha$  represents the number of dimensions with a value of  $\alpha$  in  $CFV$ ,  $D$  represents the total number of dimension,  $\theta$  represents the weight of the factor with a value of  $\alpha$ , and its value is between 0 and  $D/2$ .

In the process of calculating the commonalities, the results we get from the method mentioned above are all larger than one. However, the value of commonality between concepts should be in the range of 0 and 1. Therefore, we need to use the normalized equation to normalize  $x$  to ensure that its value is between 0 and 1. The normalized equation is as follows:

$$com = 2 * (\frac{1}{1 + k^{-x}} - 0.5) \quad (2)$$

In Equation (2),  $com$  represents the value of commonality between policies;  $x$  is the result calculated by Equation (1). Equation (2) is the normalized formula and cannot normalize the values in  $[0, 1]$  to the interval  $[0, 1]$ , where  $k = 1.01$ .

After the above calculation, we can get the commonalities between concepts. The commonalities between concept pairs are expressed in the form of triples (concept 1, concept 2, commonality).

### 3.3. Mining the Recessive Lineage Relationship

The stronger the commonalities between concepts are, the greater their impact on the lineage relationship between policies is; on the contrary, the concepts that have weaker commonness between each other have only a limited impact on the blood relationship between policies, so concepts like this can be neglected. Therefore, it is possible to filter concept pairs with weak commonality by setting a threshold of commonality.

The mining of policy recessive lineage relationship relies on similar concept pairs. The similar concept pairs extracted by commonality are used as new synonym pairs. Then, we expand the table of synonyms and calculate the lineage relationship between policies. Comparing the results with the policy lineage relationship before the introduction of these similar concepts, we can see that new lineage relationships have developed between some previously unrelated policies. Moreover, some of the original lineage relationships have been significantly enhanced. It can be inferred that the newly introduced similar concepts significantly enhance the linkages between policies, and these concepts are the recessive policy genes we want to look for. The basic process is shown in Figure 4.

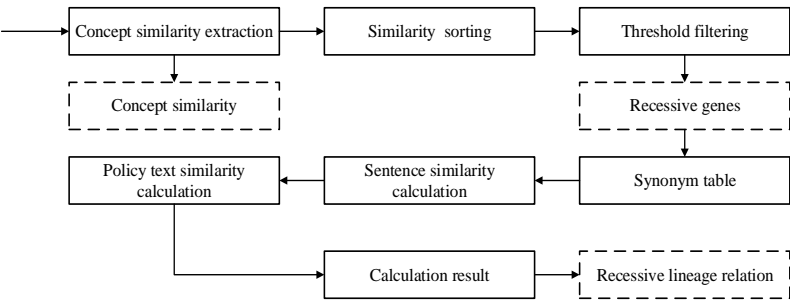


Figure 4. The process of mining the recessive lineage relationship

4. Verification and Application of Experimental Results

The experimental hardware environment is a Pentium(R) Dual-Core CPU, the dominant frequency of it is 2.50GHz, the capacity of the memory is 2G, and the capacity of the hard disk is 500G. For the development language, C # and ASP.NET are used for calculating the fitting degree of the clauses and calculating the similarity between chapters. Java is used for extracting the commonalities of conceptual factor vectors. The former runs in a Visual Studio 2010 environment, and the latter runs in an Eclipse3.2 environment. Both databases use the Microsoft Office Access 2003 database.

The experiment is divided into five stages, as shown below:

- 1st stage: preprocessing policy text.
- 2nd stage: calculating the lineage between policies for the first round.
- 3rd stage: determining the value of *BFS*, constructing the factor space and characterizing policy-specific conceptual factors through factorization. Then we calculate the common degree and set the threshold to get recessive genes.
- 4th stage: we introduce the recessive genes and then calculate the lineage between policies for the second round.
- 5th stage: we compare the two calculated results of policy’s lineage and then mining the recessive lineage relationship.

Using the above method, we have done a large number of experiments based on 1000 multi-domain policy texts. Now, we select the experiments based on policy texts of “the Law on Tendering and Bidding “and “the Law on Government Procurement” to show the process of these experiments and analyze results as follows.

4.1. Characterizing the Conceptual Factors

We disassemble 149 common concepts between policy texts of "the Law on Tendering and Bidding “and “the Law on Government Procurement”. Then, 23 basic factors are generated by iteration. By the statistical method and sorting method, we select the first 17 factors as the preferred factor set to construct the preferred factor space. The dimensions of the preferred factor space are as follows: "official", "noun", "verb", "abstract / composed of many concepts", "specific thing", "characteristic / relationship", "personal / among people ", "emerge/exist ", "established procedure ", " implied subject is a person or unofficial organization ", " negative color ", " positive color ", " money / price / agreement ", " Meet...criteria / status / existence ", " talking / external / public / announcement ", " administrative / institutional / location ", " assessment / record / material "and so on. We use the preferred factor space to decompose the factor on the unique concepts between policy texts of “the Law on Tendering and Bidding” and “the Law on Government Procurement”. The results are shown in Figure 5 and Figure 6.

	noun	verb	abstract/c omposed of multiple concepts	specific object	character /relation	personal /among people	relate d to the govern ment	emerg e/exist	establis hed process	the implicit subject is people or unofficial	negative	positive	in some condition /state/exi sting	money /price /proto col	speak /exter nal/an nounc e	administrati on/organizat ion/place	evaluate /record/ material
Save	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	1
protect	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0
promise	1	1	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0
Report	0	1	0	0	0	1	0	1	1	1	0	0	0	0	1	0	0
national	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0
this level	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0
necessary	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0
change protect	0	1	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0
standard	1	0	1	0	1	0	1	0	0	0	0	1	0	1	0	0	1
suppleme nt	1	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0

Figure 5. The result of decomposing factors on the unique concepts between policy texts of “the Law on Tendering and Bidding”

	Noun	Verb	Abstract/ compose d Of Multiple Concepts	Specific Object	Character/relati on	Personal/ AmongPe ople	Related Tothe governme nt	Emerge/exi st	Establishe d process	The implicit subject is people or unofficial organizati on	Negative	Positive	In some condition/ state/existi ng	Money/pr ice/protoc ol	Speak/exte rnal/annou nce	Administr ation/orga nization/pl ace	Evaluate /record/ma terial
Safety	0	0	0	0	1	0	0	0	0	1	0	1	1	0	0	0	0
Eight years	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
Secret	0	1	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0
Security deposit	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	1
Backup view	0	1	0	0	0	0	1	1	1	1	0	0	1	0	0	0	1
Deviate	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0
This chapter	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1
Mark	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0
Different	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Property	1	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0

Figure 6. The result of decomposing factors on the unique concepts between policy texts of “the Law on Government Procurement”

#### 4.2. Extracting the Commonalities Between Concepts

We use Formula (1) and Formula (2) to calculate the commonalities between the unique concepts of “the Law on Tendering and Bidding” and “the Law on Government Procurement”. Part of the results is shown in Table 1. We set the threshold of conceptual commonalities to 0.45, then filter 232 recessive genes to calculate the similarity between policy chapters.

Table 1. Part of the results of calculating the commonalities between policies

The unique concepts of “the Law on Tendering and Bidding”	The unique concepts of “the Law on Government Procurement”	The value of commonalities
Evaluate	Accreditation	0.63856597
Product	Stock	0.21311991
Detection	Supervision	0.50151508
Decide	Notice	0.55243859
Central Government Budget	Business license	0.57989679
Both sides	Related people	0.36886747
Leader	State organ	0.57989679

Table 2. The result of similarities between two chapters

Corresponding policy sentence	1st calculation result	Corresponding policy sentence
Sentence 1 Sentence 2	0.02677819	Sentence 1 Sentence 2
Sentence 1 Sentence 3	0.04794810	Sentence 1 Sentence 3
Sentence 1 Sentence 4	0.15097294	Sentence 1 Sentence 4
Sentence 1 Sentence 5	0.70679250	Sentence 1 Sentence 5
Sentence 2 Sentence 3	0.05554531	Sentence 2 Sentence 3
Sentence 2 Sentence 4	0.56034027	Sentence 2 Sentence 4
Sentence 2 Sentence 5	0.14395082	Sentence 2 Sentence 5
Sentence 3 Sentence 4	0.18669676	Sentence 3 Sentence 4
Sentence 3 Sentence 5	0.15578511	Sentence 3 Sentence 5
Sentence 4 Sentence 5	0.15547841	Sentence 4 Sentence 5

4.3. Result of Experiment

The result of the similarity between two chapters calculated from the experimental process is shown in Table 2. In the table, “bid x purchase y” means chapter x in “the Law on Tendering and Bidding” and chapter y in “the Law on Government Procurement”. The threshold of similarity is 0.1.

According to the result of the above experiment, we can construct a network of policy lineage relationships. Symbol A represents the policy text of “the Law on Tendering and Bidding”, and symbol B represents the policy text of “the Law on Government Procurement”. The node  $A_i$  ( $i = 1, 2 \dots 6$ ) represents Chapter  $i$  of “the Law on Tendering and Bidding” and node  $B_j$  ( $j = 1, 2 \dots 9$ ) stands for Chapter  $j$  of “the Law on Government Procurement”. If there is an edge between two nodes, it means that there is a similar relationship between the policy texts corresponding to the two nodes, and the weight labeled on the edge indicates the value of the similarity between the two policy texts. The calculated policy lineage network is shown in Figure 7(a), and the policy lineage network before the introduction of factor space theory is shown in Figure 7(b).

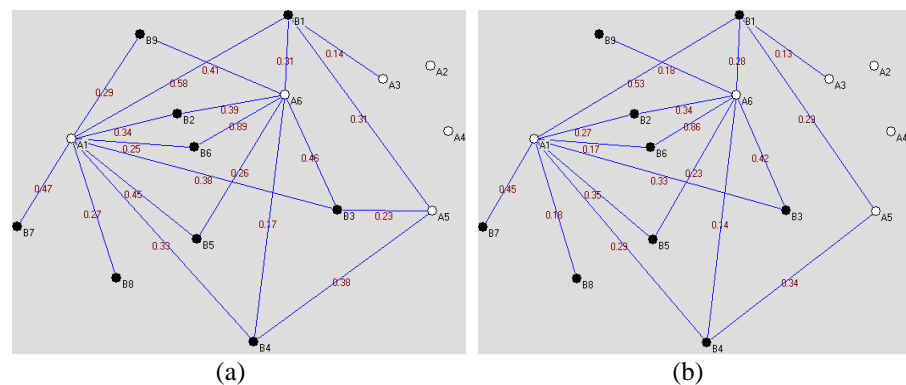


Figure 7. The policy lineage relationship network calculated in two times

Figure 7(a) and Figure 7(b) are superimposed to get Figure 8. The weight of each edge in Figure 8 is the second calculation of similarity minus the first calculation of similarity. That is, the weight of each edge in Figure 9 is equal to the weight of the corresponding edge in Figure 7(a) minus the weight of the corresponding edge in Figure 7(b). In Figure 8, when the weight of one edge is positive, this edge is indicated by solid line, which indicates that the lineage relationship is strengthened; when the weight of one edge is negative, a dotted line, it indicates that the lineage relationship is weakened. A solid line with a weight of 1 indicates the corresponding lineage relationship is strengthened and is above a threshold of 0.1. A weight of -1 in the dotted line indicates the corresponding lineage relationship is weakened and below the threshold of 0.1.

From Figure 8, we can see that the introduction of recessive genes for text similarity calculation enhances the lineage relationship. Among them, the blood relationship is greatly strengthened in “A1-B9” and “A5-B3”. For Chapter 1(A1) of “the Law on Tendering and Bidding” and Chapter 9 (B9) of “the Law on Government Procurement”, the first calculation of similarity between them is 0.0988752492338179, and the second calculation of similarity between them is 0.292782408727881. The threshold of 0.1 filters the first similarity, and the difference between the two similarities is close to 0.2, which indicates that after the introduction of the recessive gene the second time, the similarity between A1 and B9 is significantly enhanced. Similarly, the similarity between A5 and B3 is also significantly enhanced.

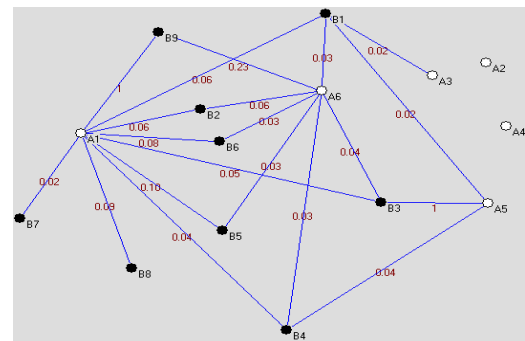


Figure 8. The superposition of Figure 7(a) and Figure 7(b)

## 5. Conclusions

Aimed at the current widespread policy inconsistency, through the study of policy, this paper reveals the reasons for inconsistencies in policy from the perspective of recessive lineage relationship and expounds the significance of mining recessive lineage relationship between policies. By introducing the theory of factor space, this paper proposes a domain-oriented method for mining recessive lineage relationships between policies. The method comprehends policy texts from the perspective of semantics and uses the concepts to represent policy genes. The cross-section of policy is the process of gene exchange and integration. Compared with previous studies that consider the word as the smallest granularity of policy, this method can identify similar recessive genes between policies, thus mining the recessive lineage relationship between policies.

Currently, it is still relatively rare to introduce the theory of factor space in the field of policy research. Although the method proposed in this paper can outperform the traditional methods of mining policy relation, anomalies can occur for individual samples, meaning the effect is not completely stable. It also needs to be further studied. Future work will further investigate the decomposition of the concept factors to avoid the interference of false recessive genes into the results of the system.

## Acknowledgements

This work is sponsored by the Postdoctoral Science Initial Funds of Heilongjiang under grant number LBHQ15031 and the Education Scientific Project of Heilongjiang under grant number GJC1215107.

## References

1. P. Buneman, W. C. Tan. 2007. "Provenance in Databases. SIGMOD." China Beijing. 1171-1173.
2. D. Fuller, L. Gauvin, M. Fournier, Y. Kestens. 2011. "Internal Consistency, Concurrent Validity, and Discriminant Validity of a Measure of Public Support for Policies for Active Living in Transportation (PAL-T) in a Population-based Sample of Adults." *Journal of Urban Health: Bulletin of the New York Academy of Medicine*. 89.2: 258-269.
3. M. Gao, C. Q. Jin, X. L. Wang, X. X. Tian, A. Y. Zhou. 2010. "Review of Database Family Management." *Journal of Computers*. 33.3: 4-27.
4. M. Ghavamzadeh, S. Mahadevan, R. Makar. 2006. "Hierarchical Multi-agent Reinforcement Learning. Auto Agents and Multi-Agent Sys." 13: 197-229.
5. H. X. Li, P. Z. Wang. 1994. "Projection of Concepts in the Factor Space." *Journal of Yantai University: session of natural and engineering*. 15-22.
6. H. X. Li, Z. L. Zhang. 1995. "Description of Fuzzy Concept in Factor Space (1)." *Journal of Kunming Engineering College*. 20.2: 16-24.
7. X. T. Peng, A. Kandel, P. Z. Wang. 1991. "Concepts, Rules and Fuzzy Reasoning." a factor space approach. *IEEE Trans. on SMC*. 21.1: 194-205.
8. P. Ren. 1992. "Factor Space and Concept Connotation Description." *Fuzzy System and math*.
9. J. Song. 2010. "The Implementing Regulations of High Operability are Preferred." *China Government Purchase*. 2.105: 30-33.
10. Y. Shang. 2009. "Game Analysis of China Central Bank's Adjustment of Interest Rates." *Business Intelligence and Financial Engineering*. 704-708.
11. W. C. Tan. 2007. "Provenance in Databases Past, Current, and Future." *Foundations and Trends in Databases*. 30.4: 22.
12. P. Wang. 2011. "Problems and Suggestions the Government Purchase System Facing." *China Government Purchase*. 2.117: 66-71.
13. X. Wu. 2008. "Analysis on the Relation between Government Purchase and Bidding." *China Government Purchase*.
14. P. Z. Wang. 1992. "Factor Space and Concept Description." *Journal of software*. 1: 30-40.
15. H.M. Zhang. 1990. "Introduction to an Expert Systems Shell-STIM." *Fuzzy Sets and Systems*. 36.1: 167-180.

**Gang Liu** received his Ph.D degree from Harbin Engineering University in China (2008) in Computer Applied Technology. He conducted research at the University of Illinois at Urbana-Champaign as a visiting scholar in Professor Jiawei Han's lab in 2005. As a member of the China Computer Federation, he has conducted g about 10 research projects such as the main researcher under the National Science and Technology Support Plan and Chinese NSFC.

**Hefei Wang** is a master's student from the School of Computer Science and Technology, Harbin Engineering University. His research interests include machine learning and natural language processing

**Honglei Zhang** received a Computer Science and Technology degree from Harbin Engineering University. Her research interests include machine learning and deep learning.