中國機械工程學刊第四十四卷第三期第 211~220 頁(民國一百一十二年)
Journal of the Chinese Society of Mechanical Engineers, Vol.44, No.3, pp211~220 (2023)

Investigating the Carbon Emission Patent Data Using the Biharmonic Spline Interpolation

Kuei-Hao Chang

Keywords: Carbon emission, biharmonic spline interpolation, irregular data points

ABSTRACT

This study reveals the patent interpolation surface for investigating the patent data ecosystem with a dataset of 3,770 patents discretized into six patent clusters. The CRISP-DM model is used to manage the patent data mining process, from data preparation to modeling and evaluating, and to explore the patent data cluster within the same framework. Four independent and three dependent factors and their corresponding three levels are identified and will be used for further data processing. The Taguchi method is used to select the orthogonal array design OA_9 (3^4) and reduce the number of cases required to investigate the patent data cluster. Using the orthogonal array design, only nine cases are performed instead of the possible 3^4, and the results are applied to structure the dataset of dependent factors. Three key dependent factors influence this patent interpolation surface: patent applications, patent assignees, and technological diversity. However, it is a difficult issue to interpolate the patent dataset using the first-order regression equation with three key dependent factors. In this case, the biharmonic interpolation method is used to interpolate irregulated patent data and create a patent interpolation surface. The patent interpolation surface is capable of interpolating the irregular data points of the dataset of six patent clusters and facilitating the analysis of the patent data ecosystem. To do so, the patent interpolation surface reveals the superposition of patent applications, patent assignees, and technological diversity to induce harmonics. As a result, harmonic traps are observed on the patent interpolation surface in conjunction with variations in patent applications, patent assignees, and technological diversity. Finally, the patent interpolation surface is divided into zones 1–4, and the harmonic traps formed are associated with a decrease in patent applications, patent assignees, and technological diversity. The levels of the three harmonic traps in zones 2–4 are 0.8, 0.8, and 0.48, respectively.

Paper Received July, 2021. Revised August, 2022. Accepted April, 2023. Author for Correspondence: Kuei-Hao Chang.

Associate Research Fellow, Innovation Headquarters, National Cheng Kung University, Taiwan 701, ROC.

INTRODUCTION

Patent data is a valuable and heterogeneous (Schröer et al., 2021) resource that is critical in driving innovation, technology licensing, and collaboration between academic institutions and industry in a variety of fields (Kim and Lee, 2015). Patent quality, intensity, technical strength, science link, science strength, and families all have an impact on the value of patent data (Magee and Yoon, 2018; Novelli, 2015). These factors comprise useful information for innovators, academic institutions, businesses, and decision-makers (Kapoor et al., 2015).

Increasing competition and patent litigation: It is true that insufficient technical information in patent documents may result in rejection by the United States Patent and Trademark Office (USPTO) (Kong et al., 2023; Chang, 2018; Somaya, 2003). As more organizations disclose their patents, competition for the legal use of patented technology will increase, making patent litigation strategies increasingly important (Burhan et al., 2017; Lissoni, 2012; Rassenfosse et al., 2013). Offensive patent litigation strategies are emerging, emphasizing the importance of patent validity and quality in preserving the benefits of industrial innovation (Wang et al., 2022).

There is a gap in the existing scholarly literature: Patent data are available from intellectual property offices, and the patent classification system distinguishes a diverse and complex range of innovation topics (Marco et al., 2019; Wittfoth, 2019). For example, PCT application data has been examined to gain a better understanding of Industry 4.0 clusters (Tsakalerou and Akhmadi, 2021). Similarly, there was an increase in global blockchain patent filings following Bitcoin and blockchain’s appearance on The Economist’s cover in 2015 (Clarke, 2022). These demonstrate the potential of patent data analysis for understanding technological trends and developments. Meanwhile, these findings highlight significant challenges in developing a patent data ecosystem, emphasizing the growing importance of addressing this issue. However, there is a noticeable gap in the existing scholarly literature, and only a thorough examination of the patent data ecosystem has the potential to bridge this gap.

Activating cross-disciplinary innovation through patent data ecosystems: The importance of patent data ecosystems is increasing as data sources become more abundant and data modeling becomes increasingly complex. The data ecosystem approach has been widely adopted in various fields, including knowledge mining, artificial intelligence (Tetteh et al., 1998), big data (Boyle et al., 2022), smart manufacturing (Pschybilla and Homann, 2020), sustainability (Yang et al., 2017), and digital economy (Neuha¨usler and Frietsch, 2020). A robust patent data ecosystem requires multiple sources of data, technological diversity (Jin et al., 2022), and complexity to meet the evolving demands of businesses and customers. To address various service demand challenges, Haak (2018) proposed a data framework to strengthen the patent data ecosystem. The patent plays a pivotal role in promoting patent value and technological innovation in various fields (Lee et al., 2022), contributing to the overall value of the data ecosystem. Detailed case studies conducted by Nylund et al. (2022) on emerging entrepreneurial biotechnology ecosystems highlight the structural, social, and ethical barriers to ecosystem development. While technologies have the potential to generate entrepreneurial ecosystems, this potential is not always realized. The Innovation Patent Index was proposed by Ponta et al. (2021) as a tool for companies to assess their innovation performance in terms of efficiency, time, diversification, quality, and internationalization. Beltagui et al. (2020) studied the evolution of patented technology in the three-dimensional printing ecosystem over the past four decades and proposed a process model of formation, growth, and destruction. Aaldering et al. (2018) visualized the structure of the business ecosystem and analyzed patent data using social network analysis. Lin et al. (2016) employed knowledge mining and data visualization techniques to quickly identify, quantify, and characterize clean energy innovation ecosystems. Gómez-Uranga et al. (2014) captured the evolution of Internet industry clusters, such as Apple, Google, Microsoft, Facebook, Amazon, and Samsung, through patent portfolios and patent litigation (Mastrogiorgio and Gilsing, 2016).

Data mining in the same framework and patent data ecosystems: Effective data mining is essential for making informed decisions, and broad frameworks play a crucial role in enabling this flow. In the realm of knowledge systems research, the DIKW framework is widely recognized for its hierarchical structure of system data, information, knowledge, and wisdom (Ackoff, 1989). This model allows for sound decision-making and wisdom (Deepu and Ravi, 2021) by utilizing data, information, and knowledge effectively to achieve the ultimate goal (Mishra, 2018). Policymakers can access related data for policy planning, evaluation projects, and service measures, allowing citizens to support government service measures by adopting a government data-sharing framework based on the DIKW model (Tungkasthan et al., 2019). Currently, studies often rely on CRISP-DM (Sharma et al., 2012), an iterative process that combines heterogeneous and diverse data into a common environmental structure to facilitate the construction of data models. This iterative process begins with domain understanding, followed by data model creation and evaluation during the data preprocessing phase. This process is repeated until the data models are deemed sufficient for the problem, after which the models are deployed in the context of the application (Artyukhov et al., 2021). Adopting a framework such as DIKW and CRISP-DM simplifies the collection, systematization, and analysis of data, facilitating the flow of knowledge for more effective solutions (Gimpel et al., 2018). However, a significant amount of data can appear in various ways in the patent data ecosystem, necessitating the establishment of a standardized process for managing data flows in this ecosystem (Harrison et al., 2014). To gain a comprehensive understanding of the data ecosystem (Heimstädt et al., 2014), it is crucial to examine the interrelationships among data users, data providers, data itself, institutions, and physical infrastructure that facilitates knowledge flows. This perspective is particularly useful for resolving complex issues, and many researchers have contributed to a deeper understanding of the data ecosystem. Mathematical models, such as the Taguchi method and the response surface method (RSM), are valuable tools for establishing relationships between design parameters and target variables. For instance, Wang et al. (2022) used RSM to predict mold aflatoxin production while examining storage conditions, such as water, temperature, and culture time. This information can be applied to designing upland rice seed storage facilities to reduce the risk of food supply shortages. Lee et al. (2022) developed a hybrid approach that combined artificial neural networks (ANNs) and RSM to optimize Antrodia cinnamomea culture conditions. ANN was used to efficiently identify the dominant factors of biomass production, while RSM explored the optimal process response. Kumar et al. (2020) constructed a reaction surface of tetraethyl lead, aircraft engine knocks, and human carcinogenic compounds to explore ways of mitigating the sources of severe health effects. Yi et al. (2019) developed an RSM-based mathematical model for the ventilation rate and gas emission rate of livestock houses. They also investigated the animal husbandry conditioning room's climate and the livestock house's gas emission rate. Xie et al. (2022) examined heat exchanger efficiency by constructing a mathematical model of the response surface. The length, arc angle, and attack angle were chosen as design parameters, with the Nusselt number and friction coefficient as the target variables. This finding has the potential to reduce the environmental effects of limited energy access and excessive energy consumption. Chen et al. (2021) found that the Taguchi method and RSM were both suitable for analyzing aerosol deposition in the lungs. The Taguchi method predicts higher aerosol deposition in aerosol flow distribution. Hou et al. (2007) used the Taguchi method, RSM, and genetic algorithm to optimize the parameters during the nanoparticle grinding process. They then used the orthogonal array design to obtain accurate reaction measurements.

A different modeling approach for unregulated patent data points: Patent data are becoming increasingly important in a variety of fields, including technology, business, and science (Legner et al., 2017; Cappa et al., 2021). However, the absence of an interpolation modeling approach in scholarly literature for unregulated patent data points presents challenges in analyzing and interpreting it. To address this issue, a standard framework is useful for managing patent datasets within the same structure. This framework is based on the widely adopted CRISP-DM standard data mining process, which facilitates the exploration of unregulated patent data points by integrating the Taguchi method and biharmonic spline interpolation. This methodology is anticipated to yield a better comprehension of unregulated patent data points and enhance the modeling process.

METHODOLOGY

This study investigates patent data x_i $(i = 1,2,3,\dots,m)$ with n parameters to create patent data clusters. To accomplish this, we create the patent data matrix, as shown in Equation (1), in which M is a matrix that includes patent data values with m rows and n columns.

$$M = \langle x_{m\times n}; i = 1,2,3,\dots,m\,; j = 1,2,3,\dots,n\rangle \quad (1)$$

To avoid singular values, M is normalized using Equation (2). In this equation, χ_{ij} represents the normalized value of the patent data element x_{ij}. Moreover, we introduced a constant, β, which ranges between 0 and 1.

$$\chi_{ij} = \frac{x_{ij}-x_{min(i)}}{x_{max(i)}-x_{min(i)}}\beta + (1-\beta) \qquad 0 < \beta < 1 \quad (2)$$

To define technological diversity for the patent data matrix χ_{ij}, we draw upon Blau's (1977) definition. Specifically, each unique subclass IPC code assigned to each patent data is denoted by IPC_u.

$$T(\chi_{ij}) = 1 - \sum_u \left(\frac{number\ of IPC_u}{Total\ number\ of\ IPC}\right)^2 \quad (3)$$

Taguchi method: In 1950, Taguchi Genichi devised the Taguchi method, a widely used approach across diverse fields. This methodology offers a distinct advantage by enabling the selection of multiple factors and their corresponding levels that influence the target value. Moreover, it allows for the observation of various factors while simulating and optimizing the process. The Taguchi method can be categorized into three significant aspects: (1) establishing factors and their levels, (2) designing the Taguchi orthogonal table, and (3) using the orthogonal array design to structure the dataset of dependent factors.

Identified factors and levels using the Taguchi method: To implement the Taguchi method effectively, the initial step involves defining the factors and levels that impact the final response parameter. In this context, factors are operational parameters that can influence output, whereas levels denote different variations in the operational factor.

Using the orthogonal array design to structure the dataset of dependent factors: The Taguchi method utilizes an orthogonal array design to obtain data and reduce the number of experimental groups required for optimization. This approach saves both time and money. The specification for the orthogonal array design is written as $OA_a(L^c)$. Here, a represents the number of cases, L represents the level combination, and c represents the maximum number of control factors that can be accommodated. Essentially, it is an array experiment $(a \times c)$. Then, an orthogonal array design is used to structure the dataset of dependent factors.

Biharmonic spline interpolation method: The method is capable of interpolating the irregular data points of the dataset of six patent clusters (Ω_1-Ω_6) and facilitating the analysis of the patent data ecosystem. The irregular surface through the n data points is expressed as follows:

$$\psi(p) = \sum_{j=1}^{n} w_j \phi_m(p, p_j) \quad (4)$$

where the data point is $p_j = (x_j, y_j)$, w_j is the weight of the data point j. $\phi_m(p, p_j)$ is the form in m dimensions of the Green's function. Furthermore, Green's functions and corresponding gradients are continuous in 1 and 2 dimensions. The irregular surface can be constructed through the linear combination of Green's functions (Sandwell,1897) [50].

CASE STUDY

To ensure precision and effectiveness in our analysis, the CRISP-DM is used as a framework for organizing the patent dataset. The Taguchi method is utilized to identify the orthogonal array design, which allows us to conduct only nine cases out of a possible (3^4). The biharmonic interpolation method is used to interpolate irregulated data points and construct a patent interpolation surface.

Understanding the response goals: In this study, the Global Patent Search System is used as our primary search tool to analyze the patent data cluster related to carbon emissions. Our search criteria employed "(carbon emission) @DE," which can generate the patent dataset. From these data, a total of 3,770 patents are identified and categorized into three groups, denoted as G_1 through G_3. Patent applications are categorized into six patent clusters (from Ω_1 to Ω_6), as illustrated in Table 1. Ω_1 comprises all 3,770 assignees, while G_2 and G_3 are divided into Ω_2 (2,914) and Ω_3 (856), Ω_4 (2,452), Ω_5 (744), and Ω_6 (574), respectively.

Table 1. Clusters and Patent Applications

Groups	Patent Applications	Clusters	Levels
G_1	1,216	Ω_1 (3,770)	L1
	1,153		L2
	1,401		L3
G_2	883	Ω_2 (2,914)	L1
	953		L2
	1,078		L3
	333	Ω_3 (856)	L1
	200		L2
	323		L3
G_3	684	Ω_4 (2,452)	L1
	845		L2
	923		L3
	339	Ω_5 (744)	L1
	172		L2
	233		L3
	193	Ω_6 (574)	L1
	136		L2
	245		L3

Table 2. Independence factors and levels for patent clusters

Independence factors	Levels		
	(2008–2012)	(2013–2017)	(2018–2022)
x_1	L1	L2	L3
x_2	L1	L2	L3
x_3	L1	L2	L3
x_4	L1	L2	L3

Identified the independent and dependent factors as well as corresponding levels: Table 2 shows the independence and dependence factors, as well as the corresponding levels that would be used for further data processing. Four factors are considered: China National Intellectual Property Administration (x_1), the European Patent Office (x_2), the USPTO (x_3), and the World Intellectual Property Organization (x_4), each of which has three different levels: L1 (2008–2012), L2 (2013–2017), and L3 (2018–2022).

Patent data clusters Ω_1–Ω_6 are analyzed by considering independent and dependent factors and corresponding levels 1–3, as shown in Figures 1–3. For example, Figure 1 presents the total number of patent applications for Ω_1 at L1. The calculations are conducted for both independent factors (x_1, x_2, x_3, x_4) and dependent factor (patent applications). The results of the calculations are as follows: x_1 (4), x_2 (138), x_3(821), x_4(253), and total patent applications (1,216). Similarly, Figure 2 presents the total patent assignees in Ω_1 at L1. The results are disclosed as follows: x_1 (3), x_2(90), x_3 (495), x_4(212), and the total patent assignees (703). Figures 3–4 present the total number of International Patent Classification (IPC) and unique subclass codes in Ω_1 at L1. According to the results, IPC and unique subclass codes include $x_1$9(6), $x_2$380 (237), $x_3$2,217 (1,467), $x_4$696 (450), and a total IPC with unique subclass 3,302 (2,160).

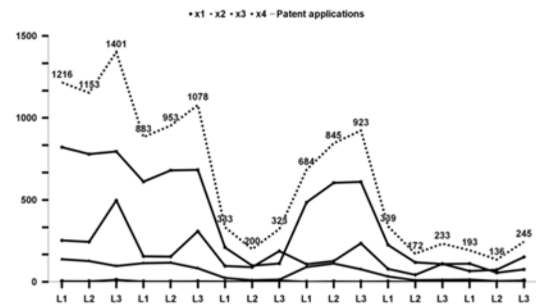


Figure 1. Patent applications of the dataset classified for independence and dependence corresponding to levels 1–3

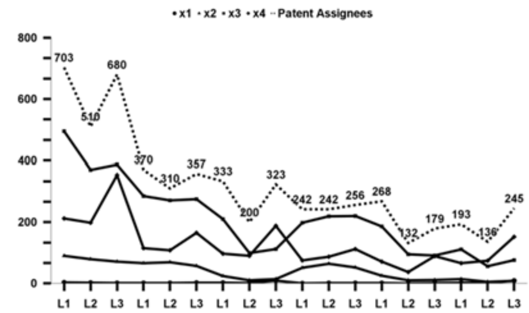


Figure 2. Patent assignees of the dataset classified for independence and dependence corresponding to levels 1–3

Using the orthogonal array design to structure the dataset of dependent factors: An insufficient interpolation node is a significant issue with biharmonic spline interpolation because it may result in either over-smoothing or discontinuity of the interpolation surface. This means that without enough interpolation nodes, the surface can become too simplified and lose important details, or it may have gaps or abrupt changes in the surface that do not accurately represent the underlying data. To address this issue, the Taguchi method and orthogonal array design can be used to improve the smoothness and continuity of interpolation nodes. However, the patent data model developed based on 3^4 cases would be prohibitively expensive for the dataset of patent clusters Ω_1–Ω_6. To improve this, we use the orthogonal array design $OA_9(3^4)$ to structure the datasets of patent applications, patent assignees, and technological diversity. The datasets for patent applications, patent assignees, and technological diversity are normalized using Equations (2) and (3), respectively. The normalized datasets are presented in Tables 4–6.

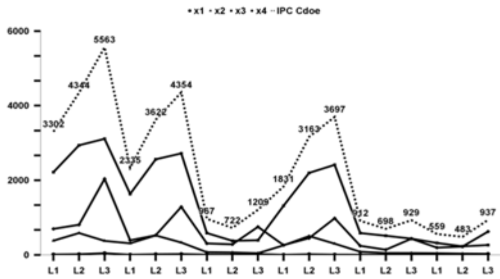


Figure 3. IPC codes of the dataset classified for independence, dependence, and clusters corresponding to levels 1–3

Enhancing modeling and facilitating comprehension: Figures 1–4 illustrate the normalized dataset of the independent factors (x_1, x_2, x_3, x_4) and dependent factors: patent applications (Y_1), patent assignees (Y_2), and technological diversity (Y_3). Regression analysis and analysis of variance (ANOVA) were used to identify significant factors and create first-order regression equations for three parameters—patent applications (Y_1), patent assignees (Y_2), and technological diversity (Y_3) However, it is challenging to observe the harmonic traps using the first-order regression equation (5). ($R^2 = 0.982$, $Adjusted\ R^2 = 0.964$).

$$Y_1 = 0.1096 + 0.6459 Y_2 + 0.6494\, Y_3 \tag{5}$$

Table 3. The ANOVA of the patent cluster model

Source	Coeff.	Std. Error	t	p*
Const.	0.1096	0.02615	4.19274	0.00
Y_2	0.6459	0.05325	12.13048	0.00
Y_3	0.6494	0.05364	12.10503	0.00
*Significant at p < 0.005				

This means that the challenge of the regression analysis was demonstrated by identifying harmonic traps in the patent interpolation surface. Nonetheless, the regression analysis indicated that three parameters (Y_1- Y_3) are capable of creating harmonic traps in the patent ecosystem through superposition combinations. In the next step, the biharmonic interpolation method is used to interpolate irregulated data points (Y_1, Y_2, Y_3) and construct a patent interpolation surface using MATLAB software. The patent interpolation surface reveals the superposition of patent applications, patent assignees, and technological diversity. The patent interpolation surface is expressed by Equation (6).

$$\psi(p) = \sum_{j=1}^{n} \quad w_j \phi_m(p, p_j)\ , \tag{6}$$

Where n is the number of data points. The patent interpolation surface is an irregular surface, which is useful for forming harmonics, as seen in Figure 5(a). In Figure 5(b), three harmonic traps are observed on the patent interpolation surface; these traps are formed in response to patent applications, patent assignees, and technological diversity.

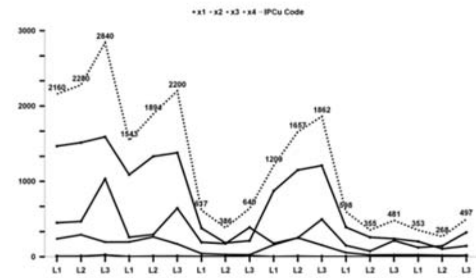
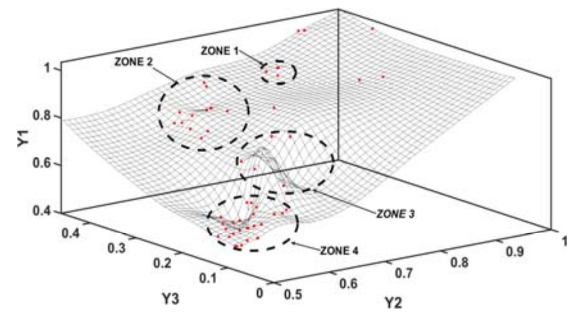
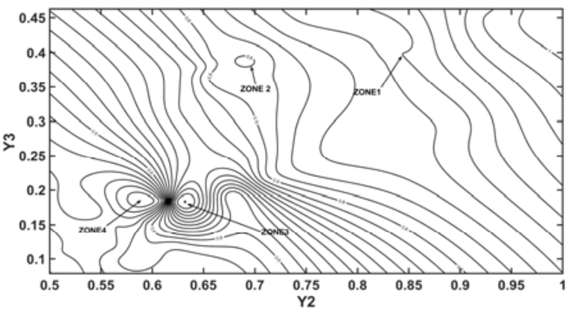


Figure 4. Unique subclass codes of the dataset classified for independence, dependence, and clusters corresponding to levels 1–3



(a) Three harmonics formed in conjunction with patent applications (Y_1), patent assignees (Y_2), and the technological diversity (Y_3)



(b) Three harmonics formed at various contour levels

Figure 5. The patent interpolation surface is characterized by three parameters: patent applications (Y_1), patent assignees (Y_2) , and technological diversity (Y_3). (a) Three harmonics formed on the patent interpolation surface. (b) Three harmonics formed at various contour levels

Table 4. Normalized datasets of patent applications based on patent clusters 1–6

The dataset of patent applications (Y_1)						
Exp.	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6
1	0.9087	0.7826	0.5742	0.7072	0.5765	0.5212
2	0.8852	0.8091	0.5242	0.7674	0.5140	0.5000
3	0.9758	0.8561	0.5678	0.7973	0.5364	0.5383
4	0.9848	0.8667	0.5663	0.8011	0.5473	0.5326
5	0.8943	0.8110	0.5314	0.7633	0.5231	0.5042
6	0.8894	0.7701	0.5674	0.7098	0.5542	0.5216
7	0.8985	0.8095	0.5371	0.7621	0.5189	0.5136
8	1.0000	0.842	0.6061	0.7636	0.5795	0.5530
9	0.8803	0.7973	0.5311	0.7473	0.5284	0.5008

Table 5. Normalized datasets of patent assignees based on patent clusters 1–6

The dataset of patent assignees (Y_2)						
Exp.	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6
1	0.9191	0.7086	0.6239	0.6182	0.5923	0.5354
2	0.8243	0.6972	0.5405	0.6473	0.5038	0.5000
3	0.9279	0.7282	0.6131	0.6561	0.5348	0.5638
4	0.9286	0.7314	0.6106	0.6555	0.5455	0.5544
5	0.8432	0.7042	0.5525	0.6410	0.5221	0.5070
6	0.8976	0.6985	0.6125	0.6271	0.5613	0.5360
7	0.8407	0.6985	0.5619	0.6403	0.5107	0.5228
8	1.0000	0.7427	0.6770	0.6504	0.5942	0.5885
9	0.8268	0.6947	0.5518	0.6327	0.5259	0.5013

Table 6. Normalized datasets of technological diversity based on patent clusters 1–6

The dataset of the technological diversity (Y_3)						
Exp.	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6
1	0.2686	0.2158	0.1221	0.1842	0.1108	0.0786
2	0.4081	0.3756	0.1231	0.3782	0.1259	0.0839
3	0.4633	0.4225	0.1855	0.3936	0.1532	0.1514
4	0.4509	0.4315	0.1919	0.3657	0.1708	0.1544
5	0.4068	0.3477	0.1195	0.3868	0.1149	0.0789
6	0.2939	0.2445	0.1171	0.2241	0.1012	0.0785
7	0.4017	0.3758	0.1180	0.3514	0.1100	0.0855
8	0.3890	0.2927	0.1928	0.3254	0.1516	0.1511
9	0.3862	0.3818	0.1212	0.3179	0.1358	0.0819

RESPONSE RESULTS AND DISCUSSIONS

To gain a better understanding of the goals of patent data modeling, we identified six distinct clusters, denoted as Ω_1–Ω_6 in Tables 1–2. This procedure enables a more thorough understanding of the data and more effective modeling.

Identifying independent and dependent factors as well as corresponding levels: A preliminary analysis of the data is required before building the regression models, which is followed by the selection of appropriate independent and dependent factors and the determination of their corresponding levels for further processing. Figures 1–4 present patent data from six identified clusters relating to four independent and three dependent factors. The results include datasets pertaining to patent applications, patent assignees, IPC code, and unique subclass code. Figure 1 shows that the total number of patent applications exceeds 10^3, including Ω_1 in levels 1–3 (1,216, 1,153, and 1,401) and Ω_2 in level 3 (1,078). Meanwhile, Figure 3 shows that the total number of IPC codes exceeds 10^3, including Ω_1 in levels 1–3 (3,302, 4,344, and 5,563), Ω_2 in levels 1–3 (2,335, 3,622, and 4,354), $\Omega_3$in level 3 (1,209), and Ω_4 in levels 1–3 (1,831, 3,163, and 3,697). Furthermore, as shown in Figure 4, there are more than 10^3 unique subclass codes, with Ω_1 in levels 1–3 (2,160, 2,280, and 2,840), Ω_2 in levels 1–3 (1,543, 1,894, and 2,200), and Ω_4 in levels 1–3 (1,209, 1,657, and 1,862). This framework is necessary to normalize irregularly scaled datasets, which streamlines data collection, systematization, and analysis by allowing the integration of all relevant data sources and analysis techniques.

Using the orthogonal array design to structure the dataset of dependent factors: The Taguchi method is used to choose the orthogonal array design OA_9 (3^4) for exploring patent data clusters Ω_1–Ω_2, as well as to reduce the number of required cases. Using this orthogonal array design, we only run 9 cases out of a possible 81.

Enhancing modeling and facilitating comprehension: SPSS Statistics 17.0 software is used to analyze patent data clusters based on the normalized dataset shown in Figures 1–4. The goal is to determine the relationship between four independent factors and three dependent factors. Table 3 presents the ANOVA results for patent applications, patent assignees, technological diversity, and patent clusters. The level of significance is set at 0.005 to ensure the reliability and reproducibility of the regression model. To do so, Equation 5 ($R^2 = 0.982$) represents the patent cluster model with three variables: patent applications (Y_1), the total number of the assignee (Y_2), and technological diversity (Y_3). However, Equation 5 is not suitable for forming an interpolation surface by interpolating irregular points. Equation 6 is the interpolation surface, which is constructed through the superposition combination of Y_2 and Y_3. In Figure 5, the patent clusters are characterized using three parameters: patent applications, patent assignees, and technological diversity. As previously stated, the patent interpolation surface would be a useful insight for interpolating irregular red data points as shown in Figure 5(a). It means that harmonic traps are observed when the patent interpolation surface is obtained using the biharmonic interpolation method to interpolate the irregular points of the patent data ecosystem. As a result, the patent interpolation surface reveals that the harmonic traps formed are associated with a decrease in patent applications, patent assignees, and technological diversity. Meanwhile, Figure 5(b) is observed that the patent interpolation surface is divided into Zones 1–4. Zone 1 has a high contribution to patent applications, patent assignees, and technological diversity, whereas Zone 4 has a low contribution to these factors. Three harmonic traps observed in Zones 2–4 have levels of 0.8, 0.8, and 0.48, as shown in Figure 5(b), respectively.

CONCLUSIONS

The patent data ecosystem can be divided into six clusters (Ω_1–Ω_6) for analyzing patent applications, patent assignees, and technological diversity. However, the patent data ecosystem contains irregular points, making it difficult to develop an interpolation surface for patent datasets. To create the interpolation surface for patent datasets, regression analysis, and ANOVA was used to identify significant factors and create first-order regression equations for patent applications (Y_1), patent assignees (Y_2), and technological diversity(Y_3). However, the data points obtained from these factors are often irregular, which presents a challenge in observing the harmonic traps using the first-order regression equation. In this version, the superposition described in this study involves combining three dependent factors to model patent data and then applying the biharmonic interpolation method to interpolate irregular data points and create an interpolation surface for patents. A known issue with biharmonic interpolation is that insufficient interpolation nodes lead to over-smoothing or discontinuity of the interpolation surface. Although the use of biharmonic interpolation is not new, the specific application of the Taguchi method and orthogonal array design to improve the interpolation of irregular patent data clusters is an innovative approach. High-order polynomial regression can fit irregular data points in a dataset, but noise can cause noticeable oscillations in the curve. Biharmonic spine interpolation is usually a better fit for irregularly spaced data points. This is because biharmonic spline interpolation uses Green's function to interpolate data points, resulting in a smooth curve that passes through all data points. This study improves issues related to irregular data points and may be considered an innovative approach to patent data modeling and analysis. As a result of variations associated with patent applications, patent assignees, and technological diversity, three harmonic traps are observed on the patent interpolation surface. Three harmonic traps located in Zones 2–4 have levels of 0.8, 0.8, and 0.48, respectively. The results show that the biharmonic interpolation method is suitable for fitting irregular patent data.

ACKNOWLEDGMENT

Financial support for this work was provided by Innovation Headquarters, National Cheng Kung University, Taiwan, R.O.C.

REFERENCES

Ackoff, R L., "From Data to Wisdom," *Journal of Applied Systems Analysis*, Vol.16, 3–9(1989).

Aaldering, L.J., Leker, J., Song, C.H., "Analyzing the impact of industry sectors on the composition of business ecosystem: A combined approach using ARM and DEMATEL," *Expert Systems with Applications*, Vol.100, 17-29 (2018).

Artyukhov, A., Volk, I., Vasylieva, T., "A multi-layered DIKW framework in education quality assurance for ensuring economic sustainability," *E3S Web of Conferences*, Vol.307, 06005 (2021).

Burhan, M., Singh, A.K., Jain, S.K., "Patents as proxy for measuring innovations: A case of changing patent filing behavior in Indian public funded research organizations," *Technological Forecasting & Social Change*, Vol.123, 181 (2017).

Beltagui, A., Rosli, A., Candi, M., "Exaptation in a digital innovation ecosystem: The disruptive impacts of 3D printing," *Research Policy*, Vol.49, 103833 (2020).

Boyle, M., Jarvis, S., Singleton, A., Jones, A., "A Position Statement: Building a Data Ecosystem in Liverpool City Region (Local) to Unlock the Value of Big Data," *Heseltine Institute for Public Policy, Practice and Place, University of Liverpool research institute*, (2022)

Chang, K.H., "Combining Concept Mapping with Entropy Information to Evaluate the Level of Patent Disclosure," *Journal of the Chinese Society of Mechanical Engineers*, Vol.39, 187-196 (2018).

Chen, W.H., Chang, C.M., Mutuku, J.K., Lam, S., Lee,W.J., "Analysis of microparticle deposition in the human lung by taguchi method and response surface methodology," *Environmental Research*, Vol.197, 110975(2021).

Cappa, F., Oriani, R., Peruffo, E., McCarthy, I.P., "Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance," *J PROD INNOV MANAG*, Vol.38, 49–67 (2021).

Clarke, N.S., Jürgens, B., Herrero-Solana, V., "Blockchain patent landscaping: An expert based methodology and search query," *World Patent Information*, Vol.61, 101964 (2022).

Deepu, T.S., Ravi, V., "A conceptual framework for supply chain digitalization using integrated systems model approach and DIKW hierarchy," *Intelligent Systems with Applications*, Vol.10–11, 200048(2021).

Gómez-Uranga, M., Miguel, J.C., Zabala-Iturriagagoitia, J.M., "Epigenetic Economic Dynamics: The evolution of big internet business ecosystems, evidence for patents," *Technovation*, Vol.34, 177-189 (2014).

Gimpel H., Hosseini, S., Huber, R., Probst, L., Röglinger, M., Faisst U. "Structuring Digital Transformation. A Framework of Action Fields and its Application at ZEISS," *Journal of Information Technology Theory and Application (JITTA)*, Vol.19, 31-54 (2018).

Hou, T.H., Su, C.H., Liu, W.L., "Parameters optimization of a nano - particle wet milling process using the Taguchi method, response surface method and genetic algorithm," *Powder Technology*, Vol. 173, 153-162 (2007).

Harrison, M.T., Pardo, T., Cook, M.E., "Creating open government ecosystems: A research and development agenda," *Future Internet*, Vol.4, 900–928(2012).

Heimstädt, M., Saunderson, F., Heath,T., "Conceptualizing Open Data ecosystems: A timeline analysis of Open Data development in the UK," *In Proc. Conference for E-Democracy and Open Governement*, Vol.12 (2014).

Haak, E., Ubacht, J., Homberg, M.V.D., "A framework for strengthening data ecosystems to serve humanitarian purposes," *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data*, Vol.85, 1–9 (2018).

Kim, J., Lee, S., "Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO,"*Technological Forecasting & Social Change*, Vol.92, 332–345(2015).

Kapoor, R., Karvonen, M., Ranaei, S., Kässi, T., "Patent portfolios of European wind industry: New insights using citation categories," *World Patent Information*, Vol.41, 4-10 (2015).

Kumar, T., Mohsin, R., Majid, Z.A., Ghafir, M.F.A., Wash, M.A., "Experimental study of the anti- knock efficiency of high-octane fuels in spark ignited aircraft engine using response surface methodology," *Applied Energy*, Vol.259, 114150 (2020).

Kong, N., Dulleck, U., Jaffe, A.B., Sun, S., Vajjala, S., "Linguistic metrics for patent disclosure: Evidence from university versus corporate patents," *Research Policy*, Vol.52, 104670 (2023).

Jin, Z., Zeng, S., Chen, H., Shi, J.J., "Creating value from diverse knowledge in megaproject innovation ecosystems," *International Journal of Project Management*, Vol.40, 646-657(2022).

Lissoni, F., "Academic patenting in Europe: An overview of recent research and new perspectives," *World Patent Information*, Vol.34, 197-20(2012).

Lin, J., Chinthavali, S., Stahl, C.D., Stahl, C., Lee, S., Shankar, M., "Ecosystem discovery: Measuring clean energy innovation ecosystems through knowledge discovery and mapping techniques," *The Electricity Journal*, Vol.29, 64-75(2016)

Legner C., Eymann T., Hess T., Matt C., Böhmann T., Drews P., Ma¨dche, A., Urbach, N., Ahlemann, F., "Digitalization. Opportunity and Challenge for the Business and Information Systems Engineering Community," *Bus Inf Syst Eng*, Vol.59, 301–308(2017).

Lee, C.W., Tao, F., Ma, Y.Y., Lin, H.L., "Development of Patent Technology Prediction Model Based on Machine Learning," *Axioms*, Vol.11, 253 (2022).

Lee, M.H., Lu, W.B., Lu, M.K., Chang, F.J.,"A hybrid of response surface methodology and artificial neural network in optimization of culture conditions of mycelia growth of Antrodia cinnamomea," *Biomass and Bioenergy*, Vol.158, 106349 (2022).

Mastrogiorgio, M., Gilsing, V., "Innovation through exaptation and its determinants: The role of technological complexity, analogy making & patent scope," *Research Policy*, Vol.45, 1419–1435 (2016).

Mishra, A., "A DIKW Architecture for Cognitive Engineering," *Procedia Computer Science*, Vol.123, 285-289 (2018).

Marco, A.C., Sarnoff, J.D., deGrazia, C.A.W., "Patent claims and patent scope," *Research Policy*, Vol.48, 103790 (2019).

Novelli, E., "An examination of the antecedents and implications of patent scope," *Research Policy*, Vol.44, 493-507(2015).

Neuha¨usler, P., Frietsch, R., "Patent families as macro level patent value indicators: Applying weights to account for market differences," *Scientometrics*, Vol.96, 1–23(2020).

Nylund, P.A., Ferràs-Hernández, X., Pareras, L., Brema, A., "The emergence of entrepreneurial ecosystems based on enabling technologies: Evidence from synthetic biology," *Journal of Business Research*, Vol.149, 728-735 (2022).

Pschybilla, T., Homann, A., "Evaluation of end-to-end process and information flow analyzes through digital transformation in mechanical engineering," *Procedia CIRP*, Vol.93, 298-303 (2020).

Ponta, L., Puliga, G., Manzini, R., "A measure of innovation performance: The Innovation Patent Index," *Management Decision*, Vol.59, 73-98 (2021)

Rassenfosse, G.De., Dernis, H., Guellec, D., Picci, L., Potterie, B.P.D.L., "The worldwide count of priority patents: A new indicator of inventive activity," *Research Policy*, Vol. 42, 720-737(2013).

Sandwell, D.T., "Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data," *Geophysica Research Letters*, Vol.14,139–142 (1987)

Somaya, D., "Strategic Determinants of Decisions not to Settle Patent Litigation," *Strategic Management Journal*, Vol.24, 17-38(2003).

Sharma, S., Osei-Bryson, K.A., Kasper, G.M. "Evaluation of an integrated Knowledge Discovery and Data Mining process model," *Expert Systems with Applications*, Vol.39, 11335-11348

(2012).

Schröer, C., Kruse, F., Gómez, J.M., “A Systematic Literature Review on Applying CRISP-DM Process Mode,” *Procedia Computer Science*, Vol.181, 526-534 (2021).

Tungkasthan, A., Poompuang, P., Intarasema, S., “Government Data Sharing Framework based on DIKW Hierarchy Mode,” *17th International Conference on ICT and Knowledge Engineering (ICT&KE)*. 2019

Tsakalerou, M., Akhmadi, S., “Agents of innovation: Clusters in Industry 4.0,” *Procedia Manufacturing*, Vol.55, 319–327(2021).

Tetteh, J., Howells, S., Metcalfe, E., Suzuki, T., “Optimisation of radial basis function neural networks using biharmonic spline interpolation,” *Chemometrics and Intelligent Laboratory Systems*, Vol.41, 17- 29(1998)

Wang, L., Zhang, Y., Yan, Y., “Offensive patent litigation strategic choice: An organizational routine perspective,” *Technovation*, Vol.122, 102663 (2022).

Wittfoth, S., “Measuring technological patent scope by semantic analysis of patent claims – An indicator for evaluating patents,” *World Patent Information*, Vol.58, 101906 (2019).

Wang, Y.B., Zhou, Y.F., Qin, Y.C., Wang, L.I., “Effect of environmental factors on the aflatoxin production by Aspergillus flavus during storage in upland rice seed using response surface methodology,” *LWT*, Vol.169, 113977 (2022).

Xie, C.G., Yan, G.X., Ma, Q., Elmasry, Y., Singh, P.K., Algelany, A.M., Makatar W.H., “Flow and heat transfer optimization of a fin-tube heat exchanger with vortex generators using Response Surface Methodology and Artificial Neural Network,” *Case Studies in Thermal Engineering*. Vol.39, 10244 (2022).

Yang,M., Vladimirova, D., Evans, S., “Creating and Capturing Value Through Sustainability,” *Research- Technology Management*, Vol.60, 30-39 (2017).

Yoon, B., Magee, C.L., “Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction,” *Technological Forecasting & Social Change*, Vol.132, 105–117(2018).

Yi, Q.Y., Wang, X.S., Zhang, G.Q., Li, H., David J., Amon, T., “Assessing effects of wind speed and wind direction on discharge coefficient of sidewall opening in a dairy building model-A numerical study,” *Computers and Electronics in Agriculture*, Vol.162, 235-245 (2019).

NOMENCLATURE

a	The number of cases
c	The maximum number of control factors
IPC	International Patent Classification
IPC_u	Each unique subclass IPC
L	Level combination
M	Patent data matrix
m	Mean experimental value
N	The unexpected output
n	The number of data point
OA	Orthogonal array design
p_j	The normalized data point
w_i	The weight of the data point
x_1~x_4	Independent factors
Y_1~ Y_3	The normalized datasets of dependent factors
χ_{ij}	Normalized patent data matrix
ψ	The general solution
ϕ_m	The form in m dimensions of the Green’s function
Ω_1–Ω_6	Patent data clusters

使用雙調和樣線插值探討碳排放專利數據

張桂豪

國立成功大學 產學創新總中心

摘要

本研究基於專利插值面探討專利數據生態系統；該數據生態系統有3,770筆專利數據可分為六個專利簇。CRISP-DM 模型用於管理專利數據挖掘過程；田口方法選擇正交陣列設計 $OA_9(3^4)$；雙調和插值法對不規則專利數據進行插值，創建專利插值面。該專利插值面揭示專利申請、專利受讓人和技術多樣性之疊加關係引發諧波效應。同時，結果表明專利插值面上觀察到隨著專利申請、專利受讓人和技術多樣性的變化而出現諧波陷阱；同時，結果也表明專利插值面被劃分為1-4區，形成的諧波陷阱與專利申請、專利受讓人和技術多樣性的減少有關；區域2-4中三個諧波陷阱之水平分別為0.8、0.8和0.48。