A Practical R&D Expenditure Statistic and Management Method Based on Spatial-Temporal Representation of Multi-factors and Data Twin Technology

Haitao Liu*, Haibo Gong, Shenjun Zheng, Yujuan Cao, Yong Hong*, Yongle Hu, Zuo Liu, and Hao Dai

***Abstract*—Nowadays, R&D Expenditure plays an important role in more and more creative activities of enterprises and other entities, especially in research activities and programs of society. However there still have a big problem that is how to collect and classify R&D Expenditure accurately. In this paper, after analyzing the restrictive collection factors on R&D Expenditure statistically, a practical scheme was provided that including R&D Expenditure Feature Vector and "Object Wood" concept were defined firstly, intelligent receipt recognizing model (IRPM), intelligent receipt persona model (IRRM) based on spatial-temporal representation of multi-factors and R&D expenditure data Twin(REDT) based on data multi relationship were developed creatively. Besides, intelligence carrier-class R&D expenditure management system (REMS) was developed based on above novel technologies and deployed on cloud with SaaS mode. For calling advantageously and updating conveniently, API standard interface and Full Stack Security Mechanism were also improved and used in REMS. Meanwhile, it was proved that REMS had better performance on assisting enterprise in collecting and using their R&D Expenditure after REMS employed by 50 industrial enterprises at first batch in practical over a period of time. There also have better economic benefits and social benefits after REMS was used by 211 enterprises in practically. Next, REMS would be utilized and tested in more scope of important entities so that the correlation technologies could be tested, iterated and optimized forward in the future. Actually, REMS is becoming R&D Expenditure industrial promoted by investor and market. Eventually, REMS would become one of the best R&D Expenditure collecting and using tools, it would not only promote R&D Expenditure increase but also become a industrial correlating with R&D Expenditure.**

***Index Terms*—R&D expenditure, statistic, spatial-temporal representation, data twin**

I. Introduction

With the raging of COVID-19, it is very difficult for coming back with the world economy. Then it is very important for open cooperating with innovation field of all nations. Innovation is an ever power which guiding development of world. Obviously innovation is one of an efficient method for exploring and solving global significant problems mutually by all nations, and against with era challenge cooperatively in order to promote peace and development of human beings.

Nowadays innovation of science and technology is very active in enterprises, including product research, method development, technical improvement, production line updating, organization ways reform and so on. Each of that is aimed at improving their competitive edge. Therefore it is an important index that R&D Expenditure of innovation which could measure the innovation of enterprise or other entity. The amount of R&D Expenditure should also explain the ability of an area or an entity. Many policies are introduced for advancing R&D Expenditure. Such as <<Measures on encouraging enterprises to increase R&D expenditure>>, <<Methods for promoting societies to increase R&D expenditures>> which were made by some local governments and so forth. All of that are aimed at increasing the R&D Expenditure and promoting innovation of society.

II. Related Works

With the progressing of science and technology, R&D Expenditure is paid more and more attention by society. Several researches around with R&D Expenditure was emerged. Some are about R&D Expenditure statistic methods or various applying scene, such as Dimitrios [1] presented an overview of the proposed R&D Information System, focused on CERIF and SDMX standards and their integration function, in their plan, they selected CERIF which has better metadata representation capability and high flexibility in forming semantic relationship as their platform's data model, and SDMX was applied as the statistical data and metadata swap standard. These two standards and their internal relations make the constructional data and metadata maintenance easily; the scheme would have provided a comparable statistics of R&D data of per country and a common survey methodology. Focusing with global temperature dampens R&D expenditure growth, Donadelli [2] provided a stochastic endogenous growth model based on empirical evidence, analyzed three theoretical channels of negative R&D expenditure which effecting of rising temperatures according to empirical evidence suggesting, proved that the government can offset these welfare costs by R&D Expenditure with 3.81% of total public expenditure. Haseeb [3] analyzed the short-term and the long-term impact

Manuscript received February 10, 2022; revised March 8, 2022; accepted June 12, 2022.

Haitao Liu and Haibo Gong are with Guangxi Artificial Intelligence and Big Data Applying Institute, Nanning, 530201, China.

Shenjun Zheng is with Hangzhou ChinaOly Technology Co., Ltd, Hangzhou, 310015, China.

Yujuan Cao is with Guangxi Academy of Social Sciences, Nanning, 530022, China.

Yong Hong is with State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, 430079, China.

Yongle Hu is with RunJian Co., Ltd, Nanning, 530022, China.

Zuo Liu is with Guangxi CAIH Smart Communication Techonology Co., Ltd., Nanning, 530000, China.

Hao Dai is with Mifpay Network Technology Co., Ltd, Nanning, 530000, China.

*Correspondence: 2019106190017@whu.edu.cn (H.L.)

doi: 10.18178/ijimt.2023.14.2.938

of economic growth, environmental pollution and energy consumption on health and R&D Expenditures, they collected data for about ten years of ASEAN countries using the auto-regressive distributed lag (ARDL) way, revealed various impacts about R&D expenditure in short-run and long-run. A panel vector autoregression model was employed for observing relationship of various factors with development R&D expenditure [4]. And subsequent R&D investment and total R&D Expenditure were contrasted. It was revealed that firms like to increase their R&D Expenditure following growth in sales products and employment labours. Of course, an effect that nearly all firms willing to increase their R&D input levels even if facing with a negative growth shock. It was studies that whether firms would cut R&D Expenditure facing with short-term earnings pressures of firm [5]. An improving smart contract method was employed tentatively for constructing supply chain model so that analyzing the complex articulation relation of big data that produced by supply chain comprehensively [6]. The two variable $EARN_{t-1} \leq 0$ and $\Delta EARN_{t-1} \leq 0$ were used for examining the lagged association between targets and R&D Expenditure. It was revealed that whatever high or low R&D input intensity firms will cut R&D Expenditure decidedly when earnings benchmark increases were missed. It was also concluded that agricultural R&D expenditure in Sub-Saharan Africa had been substantially more volatile than in other developing regions [7].

Some are about R&D Expenditure in economy. Such as Akcali [8] assumed that R&D Expenditure is very important for international competition and sustainable growth. Then it was investigated in detail that the relationship between R&D Expenditure and economic growth. About 19 countries (developing and developed) data were employed for their study with the panel data analyse. The conclusion seems that the effect of GDP is not the same parallel with the development level although developed countries are more than developing countries with the index of R&D Expenditure of GDP. Gumus [9] mainly analyzed the relationship between R&D Expenditure with economic growth, and the differs with degree of economic development. It is revealed that R&D Expenditure will be a positive and significant effecting on economic growth for all 52 countries they selected in the long run. The relationship was examined that between R&D Expenditure of business enterprise and high technology in European countries with a panel causality analysis [10]. They found R&D Expenditure of business enterprise is one of the main factors of improvement in the technological ability of high value-added production or industrial in Europe.

However, so far, there have not any system or platform which could be used by researchers and accountants together easily; not form articulation of different data which combining with financial data, program data, receipt data, knowledge rights data, man-hour data and researcher data etc; not applying by different bureau and department which have their duties with using of R&D data, such as high-tech industries evaluation, program acceptance, R&D Expenditure plus deduction, qualification obtainment and so on; and not be called by financial system, program system, tax system, statistics system and so forth with compatible interfaces. Therefore it is very necessary for developing a system or platform which could solve above problems completely.

III. Our Method

A scheme was provided that could solve above problems after R&D expenditure and other correlation factors were analyzed systematically and designed creatively. It was developed by modularity designation idea based on robot process automation (RPA), knowledge graph (KG), multi-dimensional isomerism data model (MDID) and extract-transform-load (ETL). In our scheme, robot process automation (RPA) and data twin were employed for R&D Expenditure management and business process automation of enterprises, extract-transform-load (ETL) and Multi-dimensional Isomerism data were used for the standardization, normalization and intelligence of R&D Expenditure of enterprises. Knowledge graph (KG) and mathematics modeling were applied for articulation of R&D expenditure and financial data. In conclusion, the total structure of our approach is shown on Fig. 1.

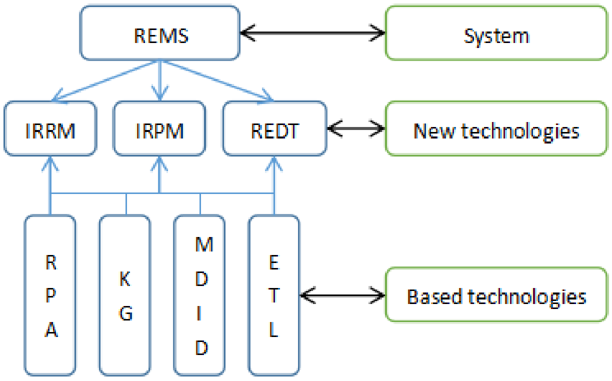

Fig. 1. The total structure of REMS.

As Fig. 1 shown, intelligence carrier-class R&D expenditure management system (REMS) was developed based on some common technologies which are robot process automation (RPA), knowledge graph (KG), multi-dimensional isomerism data model (MDID) and extract-transform-load (ETL). But the three sub-model that were developed forward based on that common technologies which are intelligent receipt recognizing model (IRPM), intelligent receipt persona model (IRRM) and R&D expenditure data twin (REDT) based on the individual ability of those technologies. The detailed introduction are as follows.

A. *Intelligent Receipt Recognizing Model (IRPM)*

Intelligent receipt recognizing model (IRPM) contains about optical character recognition (OCR) [11], intelligence classifier (IC) and automation process(AP), all receipts would be read and recognized as characters by OCR, hence the content of receipt should be obtained, which would be selected as one of the features of receipt, such as receipt number, accountant name, payment content, amount, date and so on. Then the correlating R&D contents will be selected by intelligence classifier (IC) which developed through classifying algorithm which could recognize all kinds of receipts according to the receipt content elements and correlation features produced with OCR. And all data about R&D expenditures would be selected and stored the relative memory unit by automation process (AP). So R&D Expenditure Data would be obtained after receipts were inputted on IRPM. The principle of IRPM is shown on Fig. 2.

Fig. 2. The principle of IRPM.

B. Intelligent Receipt Persona Model (IRRM)

Intelligent receipt persona model (IRRM) was developed for constructing the whole R&D Expenditure Persona Model according to R&D expenditures data obtained from IRPM, labour hours, programs which carrying out by enterprise and knowledge rights. And adding the history R&D expenditure data, removing non R&D expenditure data, Finally, R&D Expenditure Persona Model was obtained for describing spending situations of R&D expenditure of enterprise, which could help enterprise know the whole R&D Expenditure situations they spent and built the whole process management of R&D Expenditure. It is described as follows that the construction formula of R&D Expenditure Persona Model based on spatial-temporal representation of multi-factors.

Generally speaking, the common R&D Expenditure contains about labour cost (l), direct input cost (d), commissioned external research and development cost(c), depreciatory cost (de), equipment adjustment and test cost (e), other correlation cost (o), designation of new product and technology cost (dn) and amortization on intangible assets cost (a).

Vector $rdE = \{l, d, c, de, e, o, dn, a\}$ which was constructed represents R&D Expenditure Feature Vector of enterprise based on Multi-dimensional Isomerism data theory which could link every cost through program among them. Then, the vector sequence matrix would be produced for an enterprise,

$$[RDE] = \begin{bmatrix} rdE_{11} & rdE_{12} & \cdots & rdE_{1N} \\ rdE_{21} & rdE_{22} & \cdots & rdE_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ rdE_{N1} & rdE_{N2} & \cdots & rdE_{NN} \end{bmatrix} \quad (1)$$

Row of $[RDE]$ represents R&D Expenditure with time variant, column of $[RDE]$represents R&D Expenditure with program variant. So, $[RDE]$ is able to describe the R&D Expenditure of an enterprise of a period of time. And it could portray the R&D activities of enterprise of a period of time. Therefore, it equals to describe Persona on the whole R&D Expenditure of an enterprise of a period of time.

C. R&D Expenditure Data Twin (REDT)

It is worthy noting that R&D Expenditure $[RDE]$ only described the past R&D activities of enterprise, but the future R&D activities of enterprise are more known by enterprise, supervision bureau, statistical department, taxation administration, science and technology department and so on. Then, R&D expenditure data twin (REDT) was provided based on R&D Expenditure matrix $[RDE]$. It is described as follows that the specific construction method based on data multi relationship.

The R&D Expenditure vector rdE of point of time is a sequence with time variant, it is also the row of $[RDE]$. Then the Data Twin model of a sequence could be constructed as,

$$rdE(t) = \Phi[rdE(t-1), rdE(t-2), \cdots, rdE(t-d)] \quad (2)$$

Φ is a non-linear function and d is an embedded dimension among the Data Twin model.

Certainly, the R&D Expenditure Data Twin(REDT) should be constructed as,

$$[rdE(t)] = \begin{bmatrix} rdE(t)_1 \\ rdE(t)_2 \\ \vdots \\ rdE(t)_N \end{bmatrix} \quad (3)$$

And $[rdE(t)]$ could represent the whole R&D Expenditure Data Twin(REDT) of an enterprise at a time. It could predict the trend of R&D expenditures input in some extent, which is also used to estimate the amount of R&D Expenditure of an enterprise by itself or other correlation entities, and could provide R&D input orientation for enterprises or other entities which would take measures for promoting R&D activities in advance.

IV. Experiments and Applying

A. REMS Constructing

A practical system was constructed and developed according to the method which was provided by this paper for the stake of testing the performance of REMS. REMS has the abilities of multi-users using simultaneously relying on B/S structure. In REMS, intelligent receipt recognizing model (IRPM), intelligent receipt persona model (IRRM) and R&D expenditure data twin (REDT) were developed for reading a variety of receipts and utilizing in different models. All the R&D expenditures data were stored in local memory which was constructed by enterprises themselves and global memory which was constructed by science and technology department timely. Mechanism of Object Production(MOP) based on SaaS mode have been developed for producing individual API object, so that REMS would be called and integrated by other systems conveniently. Consequently, all API objects are formed a block of calling object which was named “Object Wood” figuratively based on embedded objects technology. Specifically, the top API object was constructed firstly, other objects were produced through embedded mechanism relying on the top object mutually, then a net API objects would be formed, so that the kind of API objects were various, API objects are also redundancy and API interfaces will be more standard. In addition, that could be called by enterprises and other entities for deployed on cloud with SaaS mode conveniently.

Full Stack Security Mechanism was employed for protecting data privacy security and REMS system security. And base layer applies of REMS which contains about RPA, KG, MDID and ETL. RPA should be used for inputting and writing R&D Expenditure content on system according to the kind of R&D Expenditure, so that R&D Expenditure could be collected fast. KG was employed for digging inner connections among R&D expenditures, programs, receipts and labour times. MDID should integrate different kinds and numbers of R&D expenditures, output readability uniform data. ETL was employed for making REMS adapt other system nicely such as financial system, taxation system, program management system, receipt input system and so on. It could assist REMS in extracting data, transferring data format to target format and storing data to corresponding memory conveniently. So, REMS could be completed with the function of "once reading, more times sharing" sufficiently.

B. Testing

It was recorded that the applying situations of REMS with about 50 industrial enterprises at first batch. Then the results would be used for evaluating the performance of REMS. The input intensity of R&D expenditure (IIR) is defined as follows in order to comparing with the quantity of R&D Expenditure.

$$IIR = \frac{RD_E}{T} \times 100\% \tag{4}$$

Where IIR is the Input Intensity of R&D Expenditure, RD_E is the R&D Expenditure and T is the total income of enterprise.

The R&D Expenditure input results of enterprise which not used REMS and used REMS have been analyzed in comparison comprehensively. The results are shown as following Table I.

TABLE I: THE IIR CHANGES WITH COMPARISON AGAINST IIR OBTAINED BY ORIGINAL WAYS AND IIR GAINED BY REMS

Quantity of enterprises	IIR obtained by Original ways	IIR gained by REMS(Average)	IIR Changes(Average)
48	0.7%-1.5%	2.2%-3.5%	1.5%-2%
14	1.5%-2%	2.1%-3.1%	0.6%-1.1%
3	2%-4%	2.5%-4.4%	0.5-0.4%
2	≧4%	≧4%	≧0
2	< 0.7%	< 0.7%	None

Cumulative number score (CNS) was provided and defined in order to test the performance of REMS accurately. Cumulative Number Score was defined as the quantity of enterprises which getting a certain percentage of IIR, that means the effectiveness and performance of IIR of enterprise after used REMS. It could illustrate that 48/50*100%=96% enterprises could get effective and impressive increase which is 1.5%-2% of IIR after used REMS from Table I. Of course, 0.6%-1.1% increase could also be got by about 14/50*100%=28% enterprises, 0.5-0.4% increase could also be obtained by approximate 3/50*100%=6% enterprises, and only nears 2/50*100%=4% enterprises not have obvious increase after used REMS in experiments. Remarkably, about 2/50*100%=4% enterprises have not any increase after used REMS in trials yet. The main reasons are enterprises already have standard and complete R&D Expenditure statistics works after analyzing the 4% enterprises which have not obviously been increased after used REMS. However, about 96% enterprises have progressive increase in IIR after used REMS from above analysis. That could declare the extreme effective of REMS enough.

There also have more noteworthy economic benefits and social benefits with the use of REMS. REMS was so far used by 211 industrial enterprises above state designated scale in practical used in Guangxi Zhuang Autonomous Region. REMS would be used widely by 800 enterprises in 2022, 1500 enterprises in 2023 and 3200 enterprise in 2024 according to market predicting. And the correlative value of output would be got that 4 millions yuan, 12 millions yuan and 25.6 millions yuan. Significantly, it was not only covered 80% important enterprises with their R&D Expenditure management work but also constructed a standard creatively R&D Expenditure management new approach.

From above systematic analysis and practical testing data, it would be observed that the used of REMS will make enterprises pay attention to increasing R&D expenditure and advancing R&D activities, make enterprises collect R&D expenditures as soon as possible, make the correctness of R&D expenditures of enterprises higher through cross validating of program lists, financial data, R&D spending of function of REMS. It would be advanced that R&D Expenditure input and it's level of enterprises and local region. And it would also make enterprises use REMS desirably and cultivate R&D expenditures applying industrial.

V. CONCLUSION

For better increasing R&D Expenditure and it's ratio to total income, all kinds of R&D Expenditure in enterprises were analyzed in detail. For instance, in common scene, R&D receipts and expenditures are not recognized by researchers, the statistical R&D expenditures are not sufficient, R&D expenditures cannot integrated with financial reports, there have not any measures for collecting R&D expenditures systematically and so on. All these problems are result in low of the total R&D Expenditure level, the not standard of statistical R&D Expenditure, the less utilization of R&D Expenditure. So, after analyzed all inner correlation of R&D Expenditure combining with enterprises practically, R&D Expenditure Feature Vector was defined and constructed, Intelligent Receipt Recognizing Model(IRPM), Intelligent Receipt Persona Model(IRRM) and R&D Expenditure Data Twin(REDT) were provided and developed creatively, Intelligence Carrier-Class R&D Expenditure Management System(REMS) was also developed based on B/S structure in order to apply in practical. In additional, REMS was deployed on Cloud with SaaS mode so that all users could utilize conveniently, and Full Stack Security Mechanism was provided in order to protect data and system security. The results of REMS used by 50 industrial enterprises at first batch indicate that the efficiency of the measures and the applicability of the REMS. Next, REMS would be tested in more larger scale enterprises and completed according to the testing results simultaneously. Eventually REMS will be promoted to apply and form an industrial R&D Expenditure. Actually, after 211 industrial enterprises above state designated scale in practical used, REMS is becoming R&D Expenditure industrial promoted by investor and market.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Haitao LIU conducted the whole method and wrote the paper. Haibo GONG, Shenjun ZHENG, Yujuan CAO, Hao DAI, Yongle HU, Zuo LIU and Yong HONG participated in the testing process.

Acknowledgement

This research is partly supported by the National Social Science Fund of China under Grant 21XGJ007, the Nanning Science and Technology Major Project of China under Grant 20211119, the Project of China-ASEAN AI Innovation Center in 2021 under Grant 20211118 and 20221212, the Guangxi Science and Technology Major Project of China under Grant Guike AA22068069, the Guangxi Science and Technology Key R&D Plan of China Grant Guike AB21220036 and the Open Competition Science and Technology Program of Guangxi under Grant 022JBGS014.

Reference

[1] D. C. Karaiskos, D. Xinidis, and V. Bonis, "R&D statistics information system: An interoperability tail between CERIF and SDMX," *Procedia Computer Science*, vol. 106, 2017, pp. 87–94.

[2] M. Donadelli, P. Grüning, M. Jüppner, and R. Kizys, "Global temperature, R&D expenditure, and growth," *Energy Economics*, vol. 104, 2011, pp. 1–32.

[3] M. Haseeb, S. Kot, H. I. Hussain, and K. Jermsittiparsert, "Impact of economic growth, environmental pollution, and energy consumption on health expenditure and R&D expenditure of ASEAN countries," *Energies*, 2019, vol. 12, no. 19, p. 3598.

[4] A. Coad and R. Rao, "Firm growth and R&D expenditure," *Economics of Innovation and New Technology*, 2010, no. 19, pp. 127–145.

[5] B. G. Osma and S. Young, "R&D expenditure and earnings targets," *European Accounting Review*, 2009.

[6] H. Liu, "A novel supply chain model based on smart contract, in *Proc. 2020 2nd International Electronics Communication Conference*, Singapore, 2020, pp. 115–120.

[7] G.-J. Stads and N. Beintema, "Agricultural R&D expenditure in Africa: An analysis of growth and volatility," *The European Journal of Development Research*, 2015, no. 27, pp. 391–406.

[8] B. Y. Akcali and E. Sismanoglu, "Innovation and the effect of research and development (R&D) expenditure on growth in some developing and developed countries," *Procedia - Social and Behavioral Sciences*, 2015, no. 195, pp. 768–775.

[9] E. Gumus and F. Celikay, *R&D Expenditure and Economic Growth: New Empirical Evidence*, 2015, no. 9, pp. 205–217.

[10] Ö. Karahan, *Intensity of Business Enterprise R&D Expenditure and High-Tech Specification in European Manufacturing Sector*, 2015, no. 195, pp. 806–813.

[11] P. Chirag, A. Patel, and D. Patel, "Optical character recognition by open source OCR tool tesseract: A case study," *International Journal of Computer Applications*, 2012, pp. 50–56.