iJournals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 9 Issue 3 March 2021

A Machine Learning Based Strategy to Support Decision Making of Open Innovation

Author: Yang Huang[1];Shin-Jye Lee[2]; Ching-Hsun Tseng[3]

Affiliation: Institute of Information Management, National Chiao Tung University, Taiwan[1]; Institute of Management of Technology, National Chiao Tung University, Taiwan[2];Department of Computer Science, The University of Manchester, Manchester, UK[3]

E-mail:bokxko1023@gmail.com[1];camhero@gmail.com[2]; hank131415go61@gmail.com[3]

DOI: 10.26821/IJSHRE.9.3.2021.9304

ABSTRACT

In the era of globalization, if companies solely rely on internal resources for innovation activities, it will be difficult to meet market demand and adapt to the needs of competition. Open innovation is the opening of the traditional closed-ended innovation model of enterprises, the introduction of external innovation capabilities, and the breaking of the traditional corporate boundaries. Therefore, it has become one of the important ways for modern enterprises to obtain high economic returns. However, whilst the decision makers of enterprises choose open innovation as the business model, they need to clarify their own innovation needs and spend a lot of resources and manpower to make positive decisions under the uncertain risk. Inasmuch as the rapid progression of AI technology in the industrial field, the application of AI technology in the technology management has become an increasing issue as an interdisciplinary study. In order to improve the efficiency of decision-making and streamline the operation process in an intelligent way, this work therefore applies the decision tree method based on machine learning to help decision makers of enterprises in making short-term or mid-term strategy, and the decision makers of enterprises can therefore apply information technology assistance, information integration, and data analysis to increase capacity, reduce costs, and improve decision-making quality.

Keywords: Machine learning, Decision making, Open innovation

1. INTRODUCTION

At present, the global competitive environment has gradually shifted from past price competition and responding to market demand to innovation capability based on knowledge competition. Therefore, the basis for enterprises to enhance their competitiveness is to continue technological innovation, create high value-added professional technology and master market development trends to accumulate experience and maintain competitive advantage. However, if companies rely solely on internal resources for innovation activities, it is still difficult to meet market demand and adapt to the competitive environment.

Open innovation is the opening of the traditional closed enterprise innovation model. In order to promote innovation within the organization, and intentionally and actively use the flow of internal and external technologies and ideas, it can break through the boundaries of traditional enterprises and introduce external innovation capabilities. Increase the expansion of intra-organizational innovation to market opportunities outside the organization and achieve high economic returns [1].

Our goal is to help decision makers make open innovation decisions through machine learning decision tree methods to develop short-term or medium-term strategies to reduce costs and improve decision-making efficiency.

2. RELATED WORK

2.1 Basic Concept of Decision Tree

In machine learning, the decision tree method is a predictive model that classifies quickly and the tree structure is easy to understand. Therefore, the development of related implementation software is quite common, such as SPSS, WEKA, Python, etc., and can be used to analyze rules for a wide range of applications, including engineering calculations, medical diagnosis, management science, and marketing research. Mainly for the work of classification or prediction.

Before getting started, users don't need a lot of algorithmic foundation. Decision tree learning uses a top-down recursive approach to calculate the information contribution of each type of feature. A decision tree consists of nodes and branches, root nodes and internal nodes represent a feature or attribute, and leaf nodes represent a specific classification result. The learning strategy is based on the principle of minimizing the loss function. After the decision tree is generated, the decision tree can also be pruned to prevent the decision tree from being too complicated and over-fitting.

Decision trees have many algorithms, including CART [2], ID3 [3], and C4.5 [4]. Multiple decision trees can become forests. For example, random forests are also one of the potential algorithms to date. They can adapt to complex data sets and solve the over-fitting problems often found in single decision trees.

2.2 Common Generation Algorithms

Different kinds of decision trees have different generation methods, that is, segmentation rules, which are used to measure data purity for cutting. Common generation methods include entropy, information gain, and the Gini impurity.

- In information theory, entropy represents a measure of the uncertainty of a random variable. If entropy=0, means that the data is completely homogeneous (pure), and when entropy=1, it means that the data is 50%-50%, which is impurity. When X is a discrete random variable, its entropy $H(X)$ is defined as

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \qquad (1)$$

The conditional entropy $H(Y|X)$ represents the uncertainty of the random variable Y under the condition that the random variable X is known. The conditional entropy $H(Y|X)$ of the random variable Y under the given condition of the random variable X is defined as

$$H(Y|X) = -\sum_{i=1}^{n} p_i H(Y|X = x) \qquad (2)$$

- The information gain measures the change in the purity of the data before and after the node is cut, and indicates the degree to which the uncertainty of the information of the class Y is reduced by the information of the known feature X. If the information gain of feature Aon training data set D is $g(D,A)$, then $g(D,A)$ is defined as the difference of the entropy $H(D)$ of set D and the conditional entropy $H(D|A)$ of D under the given condition of feature A.

$$g(D,A) = H(D) - H(D|A) \qquad (3)$$

In the ID3 algorithm, the larger the information gain, the better the partitioning. The decision tree algorithm is essentially to find the optimal division of each column and the order and arrangement of different column divisions. Therefore, the feature with the largest information gain is selected as the current Splitting characteristics.

Information gain is better when faced with discrete data with fewer categories, but if we are faced with continuous data such as age, height, weight, etc., or if there is no obvious category of data, it's better to use the information gain rate in this case. The information gain ratio of feature A to training data set D is defined as the ratio of its information gain to the entropy of training data set D with respect to the value of feature A, expressed as

$$g_{rate}(D,A) = \frac{g(D,A)}{H(D)} \qquad (4)$$

- In the CART algorithm, Gini impurity is a measure of the probability of randomly selecting sub-items from a data set, measuring its misclassification into other groups, or as a measure of system confusion. The smaller the Gini impurity, the higher the purity, the better the effect of classification; when the Gini impurity is 0, it means that the collection categories are consistent. For a given training data set D, assuming there are c classes, the probability that the sample belongs to the i-th classification is p_i, and the Gini impurity indicator can be calculated as follows

$$Gini(D) = \sum_{i=1}^{c} p_i(1-p_i) = 1 - \sum_{i=1}^{c} {p_i}^2 \quad (5)$$

The greater the Gini impurity, the more chaotic the data, the less easy to distinguish. When considering binary classification, the impurity-weighted sum of each result partition is calculated. Assuming that under the condition of the two classification and the feature A, then the training data set D is divided into D_1 and D_2, and the Gini impurity at this time can be expressed as

$$Gini(D|A) = \frac{D_1}{D} Gini(D_1) + \frac{D_2}{D} Gini(D_2) \quad (6)$$

3. METHOD AND EXPERIMENTS

Due to the lack of sample data, only the model architecture is described. It is expected that the experimental environment uses the Anaconda of Python 3.6, mainly using the well-developed machine learning framework sklearn module. The machine learning model set in this paper will be described in detail below, which can be divided into three steps: data collection, feature engineering, model building and training.

3.1 Data Collection

First of all, in the data source part, we must find out the data of each company that meets the experiment in this paper, to ensure the integrity of the collected data, so that each piece of data is accurate and useful. The features of the data refer to the papers written by Chen Yu-Fen, Chen Jin and others [5], and determine the feature parameters of the data set. Each feature is assigned to a category according to its characteristics. The feature categories are basic form, technical resource acquisition, manufacturing capability acquisition, internal innovation resources, and innovation performance. Table 1 shows the detailed features of each category.

TABLE I. DETAILED FEATURES

Basic form	Industrial classification, type of ownership structure, scale of enterprise
Access to technical resources	Research and development knowledge, innovation fund
Manufacturing capacity acquisition	Manufacturing capability, new product development risk, new product development costs
Internal innovation resources	Research and development personnel capability, technical workers quality, research and development equipment, production equipment, sales staff quality
Innovation performance	New product quantity per year, new product sales rate, new product development speed, innovation project success rate, patent application number

3.2 Feature Engineering

Feature engineering is a process of expressing and presenting data by performing a series of engineering processes on the original data and refining it as features for use in algorithms and models. In practice, the purpose of feature engineering is to remove impurities and redundancy in the original data, and to design more efficient features to characterize the relationship between the solved problem and the predictive model. This part requires more expertise related to the data set. Typical feature engineering includes data cleansing, feature extraction, feature selection and other processes.

3.2.1 Data Cleaning

After the data is collected, data cleaning is required at this time for simple cleaning and pre-processing. This step requires removing noise and some unnecessary features, normalizing it, or adding dummy variables to non-numeric data. If a random forest is used as a classifier, normalization or standardization is not required because the random forest divides the sample by impure function, thus simplifying the steps in modeling. At this time, if there is a missing value, we can choose to discard the data or choose a value to fill. Because the training data set should be small, if the data with the missing value is directly discarded, it is likely that the amount or dimension of the training data will decrease, and in the future, the data may be incomplete.

3.2.2 Feature Selection

Feature selection is a process of selecting features associated with a current learning task from a given set of features. Irrelevant features do not work in many cases. If all the features are included, the burden of the learning process will increase. Therefore, feature selection can appropriately remove irrelevant features and reduce the pressure of the model. There are many methods for feature selection. Because of the use of decision tree type classifiers, this paper chooses the embedded method, which combines the feature selection process with the classifier training process. Both are completed in the same optimization process. Feature selection is automatically performed during classifier training. At the end of the feature selection, the organized data is divided into training sets and testing sets to prepare for the training model.

3.3 Model Building and Training

3.3.1 Model Selection

There are many types of decision tree series. Compared with other algorithms, it is easy to understand and has high precision. Therefore, we can choose one that has the best results. Usually for the sake of convenience, a random forest [6] composed of multiple CART decision trees will be selected, which basically inherits all the advantages of the decision tree. It is the most popular decision tree series model, which can process high-dimensional data without using Make feature selection. However, if the data set is small, it is easy to overfitting. The number of trees of some algorithms such as the Gradient Boosting Tree cannot be too much, need to pay more attention when tuning.

3.3.2 Training

Model training needs to adjust parameters to get the best results. Taking random forest as an example, the following three types of features will be adjusted to improve the predictive ability of the model (using Python's conventional nomenclature).

- Max_features: Random forests allow a single decision tree to use the maximum number of features. But by increasing max_features it will slow down the algorithm. Therefore, we need the right balance and choose the best max_features.
- N_estimators: On behalf of the number of subtrees we want to create, more subtrees can make the model perform better, but at the same time make the model slower. In order to predict better and more stable results, as long as the processor can withstand, we should choose a high value as much as possible.
- Min_sample_leaf: This value limits the minimum number of samples for leaf nodes. If the number of leaf nodes is less than the number of samples, it will be pruned. The blade is the end node of the decision tree, and the smaller leaves make it easier for the model to capture the noise in the training data. The default value is 1, but in general, we should try a variety of leaf size types to find the best one.

3.3.3 Evaluation

After getting the model, how to judge whether the model is good or bad, is it over-fitting or under-fitting? The best way to do cross-validation is to evaluate the model's performance metrics by taking a portion of the original training set as a validation set. If the results of cross-validation have a strong linear relationship, we should increase the training data or reduce the feature parameters. However, an important advantage of random forests is that there is no need to cross-validate them. For each tree, about 1/3 of the samples are not involved in the generation of the kth tree, they are called the oob samples of the kth tree, and such sampling characteristics can be used for oob estimation. The oob misclassification rate is an unbiased estimate of the random forest generalization error. Its result is similar to the k-fold cross-validation that requires a large number of calculations. Therefore, cross-validation can be replaced by the oob misclassification rate in random forests.

4. INNOVATION STRATEGIES

In order to choose the most effective strategy, it is necessary to classify the various operating modes of open innovation. This work has compiled the following five open innovation strategies. With the decision tree algorithm, we can select the best strategy based on the data set.

4.1 Industry-University-Research Cooperation

The cooperation between industry, university and research institute is a kind of cooperation mode for the practical development of academic theory and the promotion of technology for enterprises [7]. It mainly refers to the cooperation between enterprises and universities and scientific research units under the risk-sharing and mutual benefit mechanism.

Its forms include cooperative research, joint establishment of entity research institutes and bases, and training of talents, sharing of equipment and technical services. Industry-university-research cooperation is conducive to establishing long-term and stable relationships between enterprises and universities or research institutes.

In this work, enterprises are provided as a leader and technology demander, providing financial support and proposing research directions, while R&D institutions and universities provide technology platforms and manpower to enterprises.

4.2 Enterprise Technology Alliance

Enterprise technology alliances describe alliances between enterprises for strategic goals [8], which means that two or more companies with independent legal status are jointly committed to the research and development of a technology or product.

Complementary or enhanced organizations that are adapted to the needs of rapid technological development and market competition are not the same as mergers and acquisitions. It can achieve complementary resources in technology, reduce the development risk and input cost of individual enterprises, and promote technological innovation, so as to be in a favorable position in market competition.

The partners are not limited, they can be competitors, non-competitors, suppliers, customers, and even companies in different fields. The forms of cooperation are also very diverse, including co-financing, joint development, and equity participation.

4.3 Technology Mergersand Acquisitions

Technology M&A is a M&A activity that aims to acquire the technical resources of the target. After the technology acquisition, the acquirer gains control of the target, that is, the internal technical knowledge is internalized through the change of ownership.

However, technology mergers and acquisitions are not necessarily the overall corporate mergers and acquisitions of the target, but focus on knowledge transfer at the target technology level. According to the purpose of the acquirer, the M&A strategy is usually carried out in order to acquire new technology or to strengthen existing technology.

Compared with other strategic methods, the use of technology mergers and acquisitions can directly control the target, so enterprises with stronger capital will choose to use this method to achieve an open innovation strategy.

4.4 Technology Purchase

Technology purchase refers to the direct purchase of technology by the entity that implements innovation through the market's outward technological innovation sources. For example, in the early stage of technology development, enterprises can quickly acquire manufacturing capabilities by purchasing exemplified technologies such as complete sets of equipment.

Packages or key equipment, proprietary technology, technology licenses, patents, technical services and consulting, including new technologies, are generally purchased.

However, due to the short communication time between the two sides of the technical transaction, the direct technical exchange between the two parties is very limited, so it is considered to be intermittent and static, and there are limitations in the introduction of technology.

4.5 Technology Outsourcing

Technology outsourcing refers to the enterprise focusing on its core technology business. On the basis of specialization, it outsources a certain technological innovation activity or one of its links to a professional technical service company. The enterprise focuses on building its core competitiveness to build a competitive advantage.

Technology outsourcing has the advantages of complementary resources, maintaining core competitive advantages, high efficiency and reducing R&D costs. Can produce huge synergies and make resources complementary [9].

However, the shortcoming is that it may lead to a serious dependence on external technology, which will reduce the R&D investment within the enterprise and weaken the R&D capability.

5. ILLUSTRATIVE EXAMPLE

The feasibility of demonstrating the model will be carried out through an example [10]. Construct a knowledge learning model by establishing a way for humans or enterprises to learn knowledge under open innovation conditions, that is, learning from open innovation. In this example, J. Yun, D. Lee, etc. constructed a causal model of the interaction model between direct and autonomous learning, and developed a mathematical model for it, and applied it to the target.

For a visual display, refer to Figure 1 below. A, B, C, D, and E represent the five strategies of open innovation. The graphics will change according to different feature design or generation methods.

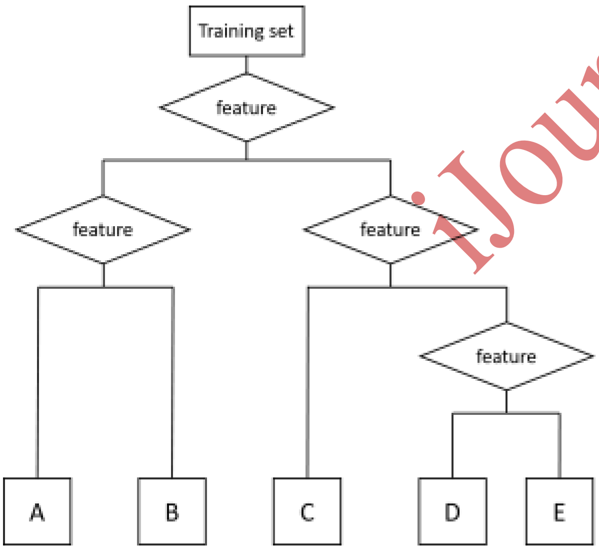

Fig 1: Decision tree concept

In addition, it is mentioned in the text that the model can be directly applied to various machine learning algorithms to improve the adaptability of the model to reality. Therefore, it should be possible to pass the decision tree algorithm mentioned in this article, but it should also be simulated under various conditions, which will make the model more complicated, and should be able to present the characteristics and elements of the data more clearly, and the final result will be an open innovation strategy can be developed for any company.

6. CONCLUSION

With the advancement of technology, data analysis tools based on machine learning technology have been widely used in various management fields. Most of the relevant research literatures in the past focused on financial technology, recommendation systems, sales forecasts, etc., but the application of innovative decision-making is lack of ink.

This work hopes to provide a complete picture of how companies can actually apply machine learning technology-based data analysis tools to open innovation decision-making processes and how these new technologies can help companies improve their effectiveness through case studies. The text shows how to apply the decision tree algorithm in machine learning to open innovation decision-making, and use examples to calculate and combine dynamic views to provide valuable empirical methods for future related research.

Managers should be good at using machine learning methods to help make the best decisions and analyze the company's innovation performance, which can provide important indicators for companies because machine learning algorithms can find important parameters to improve accuracy.

As for whether to adopt the innovative decision proposed according to the calculation results, the manager can use the analysis results as a reference, and then meet with relevant departments to decide whether to adopt this program, or find more original data to recalculate and obtain better results.

7. REFERENCE

[1] Henry Chesbrough, "Open Innovation: The New Imperative for Creating and Profiting from Technology," Harvard Business Press, 2003.

[2] Leo Breiman, Jerome Friedman, Charles J. Stone, R.A. Olshen, "Classification and Regression Trees," Taylor & Francis, 1984.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange J. R. Quinlan, "Induction of Decision Trees," Machine Learning, vol. 1, 1, March 1984, pp. 81–106.

[4] J. R. Quinlan, "C4.5: Programs for Machine Learning," Morgan Kaufmann Publishers, 1993.

[5] Chen Yu-Fen, Chen Jin, "A study on the Mechanism of Open Innovation Promoting Innovative Performance, " Science Research Management, vol. 30, 4, July 2009, pp. 1-9

[6] Leo Breiman, "Random Forests," Machine Learning, vol. 45, 1, October 2001, pp. 5-32

[7] Z. Kang, "Cooperative Vocational Education-An Inquiry," Bulletin of National Taiwan Normal University, vol. 30, June 1985, pp. 179-207.

[8] C. Huang, "Types of Strategic Alliances and Firms' Performance: Empirical Evidence from the Taiwan's Public-Listed Companies in Biotechnological Industries," Journal of ST. John's & ST. Mary's Institute of Technology, vol. 21, September 2004, pp. 115-144.

[9] Y. Gao, X. Zhou, D. Lei, "Technology Outsourcing: an Innovation Concept for Wood-based Panel Companies in China," China Wood Industry, vol. 2, 3, May 2007, pp. 45-47.

[10] J. Yun, D. Lee, H. Ahn, K. Park , T. Yigitcanlar, "Not Deep Learning but Autonomous Learning of Open Innovation for Sustainable Artificial Intelligence," Sustainability, vol. 8, 8, August 2016.