

A Knowledge Discovery Process Model Based on Reverse Engineering

Lan'An Wu^{1,2, a}

¹ National Engineering Research Center for E-learning, Central China Normal University, Wuhan 430079, China;

² Physical Science and Engineering Institute of Technology, YuLin Normal University, Yulin, 537000, China.

^awlafan@qq.com

Keywords: reverse engineering; knowledge discovery; process model; technology.

Abstract. Nowadays, the knowledge-based economy is growing rapidly. To sustain future growth, more and more well-educated people in STEM (science, technology, engineering and mathematics) are needed. But the situations of Chinese engineering education are still severe. In order to solve these problems and train outstanding engineers, this study will analyze current technological background, propose educational ideas and counterplans of the problems solving, and summarize a knowledge discovery process model based on reverse engineering. This study proposes a hypothesis-testing algorithm and illustrates the application of this process mode by a reverse engineering practice case study. It will elaborate some ideas of applying educational data mining, knowledge discovery and data mining and 3D printer to support the knowledge discovery based on reverse engineering. Our study reveals that the knowledge discovery process model based on reverse engineering is an effective way to train outstanding engineers, with the support of contemporary available technology.

1. Introduction

In this day and age, the knowledge-based economy develops rapidly. We need to motivate more young people to choose science and engineering as their future career will make our future economy competitive so more well-educated people in STEM (science, technology, engineering and mathematics) are needed. To have more students engage in science and technology, various reports suggest scientific inquiry should be included in their courses.^[1] In order to carry out the National Medium and Long-term Education Reform and Development Plan Outline (2010-2020) and the National Medium and Long-term Talent Development Plan Outline (2010-2020), and achieve the significant changes from engineering education big country to engineering education powerful country, a Plan for Educating and Training Outstanding Engineers of China has been put into practice.

It's a common phenomenon in China that doctors and masters stay at school for teaching just after they graduate from school. But as professional teachers of Chinese engineering colleges have little professional practical experience in factories, many of them are lack of practical experience. With the additional negative impact of unreasonable teaching performance evaluation index system and researching management system and short of laboratorial facilities, the problems of Chinese engineering education will still be severely for quite a long time. It mainly embodies in four ways: (1) engineering education focuses on academic research, while theories and practical operation abilities are separate from each other. (2) Talents cultivation mode is separate from real production. (3) The trained students are lack of experiences to solve actual problems and weaken in engineering practice. They are lack of thinking mode, qualities and innovation ability that an engineer should have. (4) Engineering colleges are lack of high quality practical training bases, and the technical interns hardly have access to the bases' core technologies which makes their engineering practices are something like going through the motions.

At present, an era of mass production, sharing and applying data is coming. Big data age is just around the corner. Today every walk of life has been permeated by data, and the data become crucial factors for production and promote great changes to human work, study, life and thinking. This study will propose a knowledge discovery process model based on reverse engineering In virtue of the

methods and techniques of reverse engineering and with the help of knowledge discovery and data mining, 3D printing, big data, this study will extract useful information from educational big data, discover knowledge, apply knowledge, innovate knowledge, carry out engineering practices, and then enhance learner's problem solving ability, engineering practical skills, knowledge innovative thinking and ability, finally raise the quality and level of Chinese engineering education.

2. Literature Review

2.1 Technological Background: Educational Data Mining Knowledge Discovery and Data Mining and 3D Printing.

Human society is walking to the era of big data; data may fall into three categories that are unstructured, semi-structured and structured data. Unstructured data represents 80 to 90 percent of all the current data. Educational big data is mainly made up of unstructured data. It consists of texts, images, videos, audios, anime, various kinds of files and equipment or crafts, procedures, database etc.. The core value of big data does not lie in big but to understand the value of the data by means of data mining. Generally, data mining (sometimes it is called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it to useful information. Educational data mining (EDM) is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. These information are available for teachers, learners, learner's parents, educational researchers, developers of education software system, and can predict learner's future learning behavior, study the effects of educational, promote scientific knowledge about learning and learners.

Knowledge Discovery and Data Mining (KDDM) concern the knowledge discovery process that is applied to any data source. The concept of a KDDM process model was originally discussed during the first workshop on knowledge discovery in database in 1989.^[2] KDDM process models provide the contemporary guidelines in applying data mining. The KDDM process models outline the fundamental set of steps required in data mining applications, which are typically executed iteratively with loopbacks. KDDM process models span the entire lifecycle in applying data mining, from the initial goal determination to the final deployment of the discovered knowledge.^[3]

3D printers have already been applied in many areas of everyday life. It equips pupils to understand the application and potential of this new type of technology which will be important to help them prepare for a world in which similar technologies will become increasingly commonplace, particularly in STEM contexts.^[4] 3D printing is gaining popularity in international STEM education.

2.2 Educational Ideas: Practical Education and Inquiry-Based Learning.

Originating from practice, integrating with practice and being applied in practice are the vitality of engineering theory. The engineering design practice is the core of the engineering. For engineering students, the cultivation of practical ability and creative ability needs plenty of time, and as well needs the accumulation of experiences. In a word, practical education is one of the basic characteristics of engineering education.

Moreover, students need to learn how to question the phenomena by engaging in a dialogue with the physical world. Students need to become aware of the nature of scientific ways of knowing.^[5] Curricula should be designed to help students learn how to regulate their own learning, how to continue to gain new knowledge, and how to update their existing knowledge.^[6] Inquiry learning is defined as an approach to learning that involves a process of exploring the natural or physical world, which will lead to asking questions, making discoveries, and strictly testing those discoveries in the search for new understanding. So inquiry-based learning is one of the best learning methods to STEM.

2.3 Counterplan: Reverse Engineering.

Reverse engineering is the process of extracting knowledge or design information from anything man-made and re-producing it or anything based on the extracted information.^[7] The aim of reverse engineering is to deduce the design plan and principle directly from original products, in the case that designers cannot get the necessary production information easily. It is an important way to digest and

imbibe advanced technology from original products. Practices show that reverse engineering can shorten the cycle of product design and development, speed up product updates, reduce costs and risks of new product development.

Reverse engineering is an effective way to acquire knowledge. It is widely used in the field of engineering technology such as automobile industry, aerospace industry, computer software, metallography and metalsmith, mechanical manufacture. Reverse engineering is one of the ways to legally acquire trade secret around the world. Depending on reverse engineering, many new industrialization countries catch up with developed countries' engineering technology, by means of technology import and imitating them from technique developed countries. They will innovate and exploit independently, and finally reach the world advanced level. Fast is a kind of power. Only by achieve the fast iterative of technology and product can we survive in the era of Internet. Whoever can learn new technology and ideas promptly and quickly from others can they win in the competition. The specific idea of learning and practice based on reverse engineering is to analyze the structure, function, operation and source code contain of current typical equipment, processes, system and procedures; and then to discover technical principle, design principle, technical process, knowledge, to apply and regenerate knowledge, and also master the cutting-edge technologies and skills in this field through the engineering practices.

3. Research Hypothesis

It's been a long history that people acquired knowledge through reverse engineering and big data, data mining, KDDM, 3D printer make this way and practice renascent. How to use reverse engineering to innovate learning and teaching through above-mentioned technologies?

On the basis of several years of practices and explorations, this study summarizes a knowledge discovery process model based on reverse engineering as the research hypothesis, Fig.1 shows an overview of it.

This process model has some similarities with KDDM process model. They all have complex iterative nature, and the partial steps are almost the same. Reverse engineering accepts failures and thinks creativity and innovation can be nurtured by learning from mistakes and failures. It's an iterative and long-term process. By piling up small successes to great victories, success can be achieved via repeated failures and introspection. Understanding it is very crucial to achieve success in reverse engineering practices. Double sided arrows in Fig.1 indicate that those tasks need to be carried out iteratively.

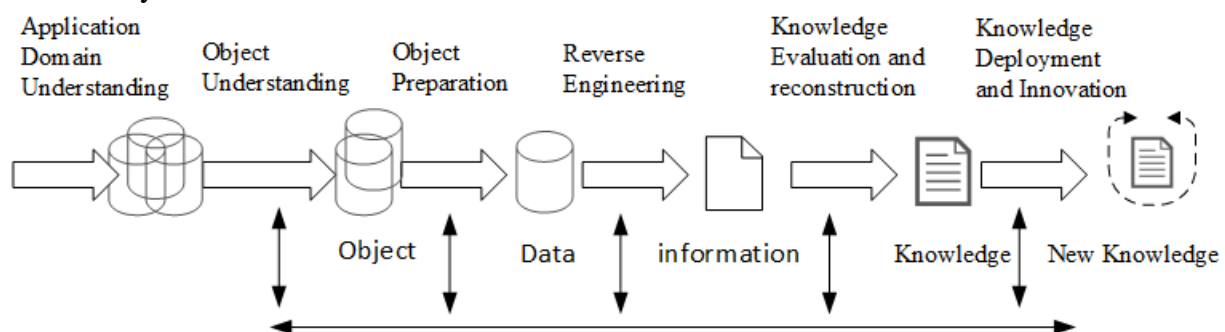


Fig. 1 the Knowledge Discovery Process Model Base on Reverse Engineering

The details of each step of this process model are as follows steps:

(1)Application domain understanding. It means understand learners' characteristics and needs, subject features, curriculum objectives, learning and teaching objectives, and converts them to the definition of reverse engineering problem. (2)Object understanding. With the support of educational big data, this step will analyze the accessibility and availability of objects that contain education information, such as texts, images, videos, audios, animation, various kinds of files and equipment or crafts, procedures, databases. (3)Object preparation. Select the object of reverse engineering, translate object to data if necessary. (4)Reverse engineering. Select reverse methods, such as means-ends

analysis, hypothesis-testing, decompiling, analogy, conclude, synthesize, spanning tree algorithm, association rule algorithm etc., apply reverse engineering to extract useful information and discover knowledge. (5) Knowledge evaluation and reconstruction. Interpreting the results, assessing the impact, novelty and interestingness of the knowledge discovered, and reconstruct the knowledge discovered. There are two typical evaluation methods: one is evaluated by domain experts, the other is evaluated by learner through the practice of forward engineering. As carrying out engineering practice can improve learners' problem solving ability, engineering practice skills, knowledge innovation thinking and ability, so in this step, it advocates practical education and inquiry-based learning. (6) Knowledge deployment and innovation. This step involves presenting the discovered knowledge in a learner-oriented way, performing deployment, engineering practice and innovation, and writing final technical documentation.

4. Methods and Process

There are many engineering disciplines. Every engineering discipline has its own discipline characteristics and laws so their specific methods of reverse engineering are not the same. In a word, reverse engineering is more than just a one-trick pony. Following their own laws, learner should look for an actual needed way to apply reverse engineering. In most cases, the reverse engineering is a practice something likes “to wade the river by groping for the stones”. As shown in the fig.1, it needs verifying the feasibility of prior solution iteratively.

Based on the process model of fig.1 and by case study verified model, this study will discuss and summarize the major features of the discovered knowledge based on process model.

4.1 The Case Study: reverse engineering MOOC micro-video webpages.

When studying in the MOOC platform, such as Coursera, XuetangX, learners find that they can accurately locate micro-video timeline by clicking on the subtitle of micro-video on webpage, as shown in Fig.2. In order to learn and master this technology, this case study will apply reverse engineering to these webpages by way of hypothesis-testing. Their algorithms are as follows:

- Assume $W(t)$ as the ultimate target code of this case, that is :user can accurate locate micro-video timeline by clicking on the subtitle of micro-video on the webpage.
- By the way of “view source code”, copy all of the HTML codes on the webpage, paste them to a Text Document and save in .html format, take all of the HTML codes as $W(t)$.
- Reduce the gradient of $W(t)$, that is $W(t-1)=W(t)-\Delta W(t)$, and $\Delta W(t)$ is a bit of code of $W(t)$. If webpage testing effect $W(t)=W(t-1)$, it means that the gradient reduction of $W(t)$ is successful, then $W(t)=W(t-1)$ and loop current step.
- Until there is no $\Delta W(t)$ that makes $W(t)=W(t-1)$, $W(t)$ will be the ultimate target code of this case, and then knowledge has been discovered, the process of reverse engineering is completed. According to the steps of Fig.1, concrete steps of this case are as follows:

(1) Study and understand the basic knowledge of web design, JavaScript, HTML, javascript.

(2) Preliminary recognize and determine the reverse objects, that is the webpages of Coursera or XuetangX, with the function that learners can accurately locate micro-video timeline by clicking on the subtitle of micro-video on webpage.

(3) Select a certain typical webpage of them as the reverse object. By means of view source code, copy its HTML code, save in Text Document and save in *.HTML format; download the files that are in *.JPG, *.JS, *.CSS, *.MP4 format on the webpage and then save them to respective file directory according to their HTML code. Acquire reference information from Internet, modify garbled HTML code and restore correct HTML code.

(4) Assume all of the HTML codes as $W(t)$, in accordance with above-mentioned hypothesis-testing algorithm, remove $\Delta W(t)$ step by step, until there is no $\Delta W(t)$ that makes $W(t)=W(t-1)$, finally get $W(t)$ as shown in Fig.3. It's the ultimate target code of this case, it consists of four key elements ,they are the three *.JS type files, one *.CSS type file, one *.SRT file, one *.MP4 type file.

(5) Analyze, interpret and study the HTML code in Fig.3. Three *.JS type files and one *.SRT type file are used to locate video timeline, the *.CSS type file is used for webpage layout, the *.MP4 type file is the played video, the *.SRT type file is the subtitle of the played video.

(6) Integrate the HTML code of Fig.3 to PHP technology, this study works out one distribution platform for mini-type micro-lecture, it can accurate locate micro-video timeline by clicking on the subtitle of micro-video on the webpage, and then the technological integrated innovation has been achieved.

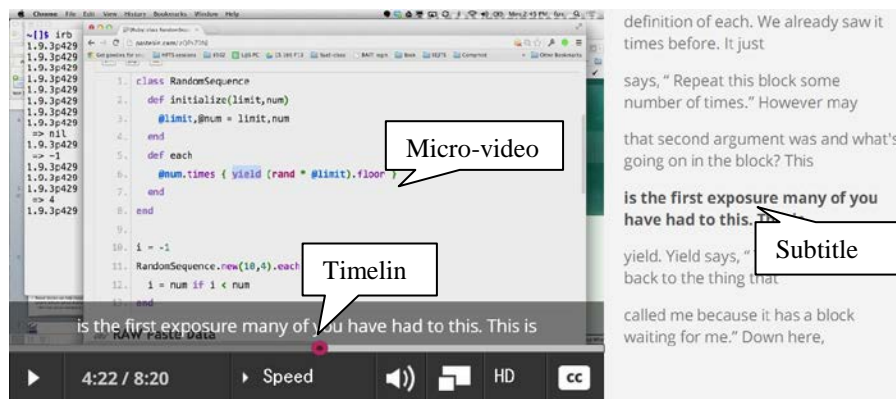


Fig. 2 Accurate Locate Micro-video Timeline by Click on the Subtitle

```
<head>
<script type="text/javascript" src="js/lms-main_vendor.5d2cd80e9461.js" charset="utf-8"></script>
<link href="css/lms-style-course.b57ec0f63cbd.css" rel="stylesheet" type="text/css" />
</head>
<body>
<h2>video title </h2>
<script type="text/javascript" src="js/videoplayer.754b1805a07d.js"></script>
<div id="T-video">
<video width="640" height="360" controls="controls">
<source type="video/mp4" src="videoname.mp4"/>
<track kind="subtitles" src="Subtitlename.srt" srclang="ch" label="ZH"/>
</video>
</div>
<script type="text/javascript" src="js/lms-modules.3b4120ae1703.js" charset="utf-8"></script>
</body>
</html>
```

Fig. 3 Ultimate Target Code of the Case

4.2 Other cases study overview.

In many engineering study, learners hardly have any chance to carry out engineering practices due to the laggard or shortage of necessary experimental equipment. At present, 3D printers can print out almost any shapes or geometrical shapes, and are produced from 3D models or other electronic data sources. The 3D printer project offers an opportunity for schools to explore innovative ways of teaching STEM subjects, stimulating pupils' interest and enriching the curriculum. There is huge potential for them to be used within a range of STEM subjects. In the field of designing physical prototype, such as automobile industry, aerospace industry, mechanical manufacture, learners can carry out forward engineering after reverse engineering tasks are completed, taking the advantages of 3D printer. 3D printer can help learners verify knowledge that are learned from classes by engineering practice, and make knowledge visualization. Accompanied with the sense of achievement, it will further arouse learners to carry out reverse engineering practices and inquiry-based learning activities, apply knowledge and innovate knowledge in practice, thus enhance learner's engineering practical ability, innovation ability and engineer thinking mode.

The reverse engineering for traditional machinery manufacture, automobile industry, aerospace industries needs to be accompanied with plenty of infrastructure, generating device, laboratory, factories, etc. and plenty of fund investment as well. Common engineering colleges hardly can bear the heavy responsibilities for reverse engineering of large-scale, high-level facilities and technology. For example, in some of the famous engineering universities in China, students of warship major can only use the ex-service destroyers, submarine as their object of teaching, experimenting and researching. In contrast, the reverse engineering with the support of big data, knowledge discovery, data excavation, and 3D print is more faster, cost and time efficient. And the possibility to popularize is high. Big data will make the learners get the world's most advanced technique and technology,

references and technical documentation, even the open-source hardware and 3D model easily. For example, using the open-source hardware and information, editable Triangular Mesh .STL files provided on one of American's creative sharing sites, learners can produce a four-axis flying machine just by below steps: components purchase, designing the center, designing the wing, printing, wires connecting, propellers assembling ,flying testing etc. So, with the help of big data, knowledge discovery, data excavation, learners can apply reverse engineering to the fields of automobile and aerospace industries, general technical and equipment of engineering, metallurgy, machinery manufacture, individual consumption etc. Then with the support of 3D print, we can produce related components and products, test and apply, innovate the knowledge. Thus enhance the abilities of engineering practices, knowledge application and innovation for engineering teacher and students.

Moreover, by means of decompiling, synthetically and usage of Sothink SWF Decompiler and Adobe Flash, learners can carry out reverse engineering of some first-class flash e-learning courses of corporate training in domestic and learn how to use Adobe Flash and Action Script 3.0 to create anime micro-lecture video, how to use domain limitation to prevent unauthorized access. Besides, by applying reverse engineering to vast MOOCs learning videos by means of analogy, induction, synthetic method and depending on the needs of different users, learners can write multiple types script formats of micro-lecture or courseware creatively, thus achieve the original innovation of method and technique.

What is more, in big data era, learners can obtain educational big data form MOOCs, online education platform, education cloud platform, various academic journals database such as Chinese National Knowledge Infrastructure and Science Direct through the Internet. Learner could take these data as educational big data and apply KDDM technology, such as spanning tree algorithm and association rules algorithm, to support reverse engineering and innovate education through technology.

5. Discussions and Conclusions

Several years of practices and researches show that the knowledge discovery process model based on reverse engineering is effective for many engineering disciplines. Besides, this study finds out some laws of discovering knowledge based on reverse engineering.

(1)Reverse engineering is not an imitation process, discover and apply knowledge are its goal and innovate knowledge is its soul. The nature of reverse engineering process is “prototype recognition, prototype reproductivity, prototype beyond”. It is not enough just to study and master the advanced technology contained in the object of reversed engineering, otherwise, we will fall behind others forever. After acquire knowledge from reverse engineering, learners need to accomplish the goal of re-innovation based on assimilation and absorption of the imported technology, integrated innovation, original innovation Only master the best advanced core technology can learners become indispensable excellent engineer or engineering technologist.

(2)Reverse engineering requires learner master certain foundation of research and development. That requires teachers have higher research and development ability, research and innovation ability. Before the activity of reverse engineering, teachers must ensure that learners have already master relevant basic knowledge, as shown in Fig.1, the first step is application domain understanding. The higher of learner's research and development ability, the higher of their reverse engineering ability it is.

(3)The objects of reverse engineering must be chosen carefully. It is not that the objects of reverse engineering should be the state-of-the-art technologies, while the most suitable object is the best for reverse engineering. In engineering education classes, as a case study and discuss object, the object of reverse engineering must be representativeness, and its reverse engineering process can't too complex to carry out. If the reverse engineering practice is too difficult for learner to comprehend, learners may lose interest and confidence of reverse engineering practice. Teachers and learners should follow certain teaching laws, such as step by step learning, from easy to difficult learning etc..

(4)Reverse engineering is one kind of high difficulty and time-consuming exploratory process, however, collaborative can do it much better. Reverse engineering needs verify the pre-plan over and over again. Query and puzzle are the natural power source of exploratory process. Learners must have patience, faith, perseverance and fighting will. Our study shows that collaborative learning can make reverse engineering more efficient, and make the promoting effect of learner's creativity and innovation ability more obvious.

(5)Reverse exploration is a way of creative thinking, reverse engineering is a method of acquiring knowledge and a way of problem solving in engineering practice. Knowledge exploration, knowledge acquisition and engineering practice run through whole process of reverse engineering.

In the era of big data, the supports of data mining, KDDM, 3D printing, the knowledge discovery process model based on reverse engineering can make learners acquire the core and frontier technology of engineering domain, keep the advanced of engineering curriculum content, accomplish fast iterative of engineering knowledge and seamless link classroom with social needs. In the process of reverse engineering, learners can take part in engineering practice and discover knowledge and will confront with queries or puzzles. And during the process of apply knowledge in a creative way, they can enhance their problem solving ability and engineering practice ability, enhance their achievement motivation and inquiry awareness. This will make learning really originate from practice; combine with practice, applied in practice. To a certain extent, this process model can solve some practical education problems of Chinese engineering education, such as focuses just on academic research and neglect practice, theory and ability of practical operation are separate from each other, lack of experience to solve practical problems, weaken in engineering practice. In a word, with the support of contemporary available technology, the knowledge discovery process model based on reverse engineering is an effective way to train outstanding engineers.

6. Limitations and Suggestions

Due to limited conditions and difference of academic discipline, this study has not yet applied and popularized this knowledge discovery process model based on reverse engineering to every engineering discipline.

To the reference of the knowledge discovery process model based on reverse engineering, this study suggests researchers and practitioners of engineering education explore concrete methods and techniques of reverse engineering, apply them in engineering education, and train outstanding engineers with survival skills in the 21st century in virtue of all kinds of available massive data and rich resources in big data era.

7. References

- [1]. Sten Govaerts, Yiwei Cao, Andrii Vozniuk, et al. Towards an Online Lab Portal for Inquiry-Based STEM Learning at School[C]// ICWL. 2013:244-253.
- [2]. Klösgen, Willi, Zytow, Jan M. Handbook of data mining and knowledge discovery [J]. Kybernetes, 2003, 33(7):372-374.
- [3]. Athauda R, Tissera M, Fernando C. Data Mining Applications: Promise and Challenges [M]// Data Mining and Knowledge Discovery in Real Life Applications. 2009.
- [4]. Education D F. 3D printers in schools: uses in the curriculum: enriching the teaching of STEM and design subjects [J]. Department for Education, 2013.
- [5]. In M, J. Inquiring into inquiry learning and teaching in science [J]. American Association for the Advancement of Science, 2000.
- [6]. De J T. Computer simulations. Technological advances in inquiry learning [J]. Science, 2006, 312(5773):532-3.
- [7]. Eilam E. Reversing: Secrets of Reverse Engineering [J]. And Sons Isbn, 2005.