Research on Big Data Impact and Implementation Path of Education Industry

Chen Wei¹, Shan Yuting²

(¹ Guangdong Vocational College of Science and Technology, ² Zhuhai Gree Electrical Appliances Co., Ltd; Zhuhai, Guangdong, 519090)

Abstract

This paper mainly focuses on what kind of data should be collected, what should be integrated and analyzed in the era of big data, and what kind of value data can bring to students, parents, teachers, schools, institutions and governments. Then it explains in detail how the education industry implements big data technology. Finally, it is imperative to summarize the implementation of big data in the education industry.

Key words: Implementation path ; Big data ; Education

Investigación sobre el impacto de Big Data y la ruta de implementación de la industria de la educación

Resumen

Este documento se centra principalmente en qué tipo de datos se deben recopilar, qué se debe integrar y analizar en la era de los grandes datos y qué tipo de datos de valor pueden aportar a los estudiantes, padres, maestros, escuelas, instituciones y gobiernos. Luego explica en detalle cómo la industria de la educación implementa la tecnología de big data. Finalmente, es imperativo resumir la implementación de big data en la industria de la educación.

Palabras clave: Ruta de implementación ; Big data ; Educación

Introduction

Big data is becoming more and more popular. Everyone is thinking about how to apply big data to their own industry and make profits. In the education industry, there are also many teachers and educators concerned about the application of big data in the education industry. In the past, education was mainly a combination of teaching, training and examinations. It used to measure students’ mastery of knowledge by examinations. It mainly focused on the evaluation of learning outcomes, but ignored the analysis of learning process. There was no intervention in the process. Students, parents, teachers and schools could only judge students’ information and evaluation by final examinations, and there was no process-level analysis.

This paper mainly focuses on what kind of data the education industry should collect, integrate and analyze, and what kind of value the data can bring to students, parents, teachers, schools, institutions and the government in the era of big data.

1 Big Data Brings Five Impacts and Changes to the Education Industry

Education concerns the national economy and people’s livelihood. Big data has great potential in reforming education. Big data is coming quietly, and data that could not be collected and analyzed in the past is given the possibility under the present technology. Who can use big data more profoundly, who can take the lead in
the development of their own industry, the same is true in the field of education. I think big data will bring influence and reform to the education industry from the following aspects:

1.1 Big data make students learn more efficiently and have more fun.

The traditional learning mode is generally: preview before class, listening in class, doing exercises after class, plus review and consolidation. The most effective tool to measure the mastery of knowledge points is the correct rate of problem-solving. The market is full of a large number of test books, a variety of, often dazzling students. Although there are many high-quality question sets which are deeply analyzed by famous teachers and experts. However, when the students do not understand a certain knowledge point thoroughly, there are errors in answering questions, and through the analysis of the answers to understand the reasons for the errors, the students may not know whether they really grasp this knowledge point. When we change the angle and scene, the students will make mistakes again.

Intelligent recommendation system based on big data can quickly recommend the same type of topics for students. Students can quickly verify their mastery and correct their thinking deviation in time, so as to make continuous progress. The map of students' knowledge recorded by big data can form various visual charts, such as radar charts and so on. Based on this, students can easily solve the so-called "leak detection and filling" to save learning time and improve learning efficiency. In the traditional mode, we usually prepare the wrong questions, put the questions we usually do wrong in it, and take them out to do it every time. But we don't know what others often do wrong, what they often do wrong, and how to classify the wrong questions. But based on data, we can easily solve these problems and choose more accurate questions.

Based on the electronic answering platform, students can also introduce some incentives in the process of learning, such as scorecard, if students answer correctly, there will be points reward, if they answer incorrectly, there will be No. Because they can get real-time feedback and know their own good and bad in time, because if they want to get a higher score, students will tend to do the right topic, rather than do more, and gradually develop good learning habits, learning easier and more efficient.

1.2 Big data make students learn more efficiently and have more fun.

The traditional learning mode is generally: preview before class, listening in class, doing exercises after class, plus review and consolidation. The most effective tool to measure the mastery of knowledge points is the correct rate of problem-solving. The market is full of a large number of test books, a variety of, often dazzling students. Although there are many high-quality question sets which are deeply analyzed by famous teachers and experts. However, when the students do not understand a certain knowledge point thoroughly, there are errors in answering questions, and through the analysis of the answers to understand the reasons for the errors, the students may not know whether they really grasp this knowledge point. When we change the angle and scene, the students will make mistakes again.

Intelligent recommendation system based on big data can quickly recommend the same type of topics for students. Students can quickly verify their mastery and correct their thinking deviation in time, so as to make continuous progress. The map of students' knowledge recorded by big data can form various visual charts, such as radar charts and so on. Based on this, students can easily solve the so-called "leak detection and filling" to save learning time and improve learning efficiency. In the traditional mode, we usually prepare the wrong questions, put the questions we usually do wrong in it, and take them out to do it every time. But we don't know what others often do wrong, what they often do wrong, and how to classify the wrong questions. But based on data, we can easily solve these problems and choose more accurate questions.

Based on the electronic answering platform, students can also introduce some incentives in the process of learning, such as scorecard, if students answer correctly, there will be points reward, if they answer incorrectly, there will be No. Because they can get real-time feedback and know their
own good and bad in time, because if they want to get a higher score, students will tend to do the right topic, rather than do more, and gradually develop good learning habits, learning easier and more efficient.

1.3 Big data make students learn more efficiently and have more fun.

The traditional learning mode is generally: preview before class, listening in class, doing exercises after class, plus review and consolidation. The most effective tool to measure the mastery of knowledge points is the correct rate of problem-solving. The market is full of a large number of test books, a variety of, often dazzling students. Although there are many high-quality question sets which are deeply analyzed by famous teachers and experts. However, when the students do not understand a certain knowledge point thoroughly, there are errors in answering questions, and through the analysis of the answers to understand the reasons for the errors, the students may not know whether they really grasp this knowledge point. When we change the angle and scene, the students will make mistakes again.

Intelligent recommendation system based on big data can quickly recommend the same type of topics for students. Students can quickly verify their mastery and correct their thinking deviation in time, so as to make continuous progress. The map of students’ knowledge recorded by big data can form various visual charts, such as radar charts and so on. Based on this, students can easily solve the so-called "leak detection and filling" to save learning time and improve learning efficiency. In the traditional mode, we usually prepare the wrong questions, put the questions we usually do wrong in it, and take them out to do it every time. But we don't know what others often do wrong, what they often do wrong, and how to classify the wrong questions. But based on data, we can easily solve these problems and choose more accurate questions.

Based on the electronic answering platform, students can also introduce some incentives in the process of learning, such as scorecard, if students answer correctly, there will be points reward, if they answer incorrectly, there will be No. Because they can get real-time feedback and know their own good and bad in time, because if they want to get a higher score, students will tend to do the right topic, rather than do more, and gradually develop good learning habits, learning easier and more efficient.

1.4 Big data makes it possible for schools to teach students in accordance with their aptitude.

Because of the accumulation of data, a large amount of data is enough to form student user portraits, interest maps, subject preferences and learning trajectories. With data mining model, students can be classified and suitable teachers can be arranged for different categories of students, so it is possible to teach students in accordance with their aptitude.

1.5 Big data enables organizations to continuously improve their products and service quality.

When the organization is large enough, the difference of resources may affect the overall quality, and the disadvantage affects the overall development. The above bottlenecks can be solved by establishing standardization system mechanism. However, without an effective data monitoring and quantification system, standardization may become formalized. Therefore, effective data monitoring can promote the implementation of institutional standardization. Data monitoring system can help organizations to find and locate problems quickly and effectively. For example, most teachers do not use a certain courseware. Is there something wrong with the courseware itself, or is there another reason? The relevant person in charge of the analysis. For example, in some lectures, most teachers do not use them in the established order, which means that the curriculum arrangement should be adjusted. Because of the large data resources, organizations can also make market positioning analysis based on the data, so as to determine the location problem. For example, if we clearly know how many people are in each primary school in Beijing, where 90% of the institutional students are studying, where only 10% may be, we can conclude that 90% of the occupancy has been relatively saturated. For these places, we need to stabilize the occupancy at this.
time. For 10% of the places, we need to make a strategy analysis of recruiting new students, such as adding new teaching points, opening activity lectures, etc.

2. The implementation path of education big data.

The whole life cycle of implementing big tree data technology often goes through five stages, including data acquisition, data storage, data processing, data analysis and data presentation. The implementation of big data in education also follows this rule. The following will explain in detail how to implement big data in education industry.

2.1 Data acquisition

“Smart women can’t cook without rice”, no matter in that industry, if there is no data, talking about big data is just empty talk. Many people think that data acquisition is only a technical problem. They think that building a log system can achieve data acquisition, but in fact it is not.

Victor Mike, author of The Future of Peer Learning and Education with Big Data, argues that big data has three core elements in improving learning: feedback, personalization and probability prediction. Feedback is to establish a perfect feedback system, implement feedback for students' learning process, improve students' learning situation through feedback, and eventually form a virtuous circle to promote students' learning. Personalization is the accumulation of learning data based on individuals. Each person's situation is different, and the accumulated learning data are different. Using modern information means, we can understand the individual learning characteristics and provide individualized teaching assistance support for each person's learning characteristics.

Probability prediction is the method of large data to achieve the above purpose. The prediction based on large data can not be 100% accurate. The reference provided by large data is only a probability. The decision maker should consider what kind of countermeasures to take when the probability reaches.

First of all, we should avoid individualization and probability, say feedback, and establish a perfect feedback system. Whether the feedback system is good or not determines whether users like to use it or not, and whether they like to use it or not. The user's use or not determines the analyzability of the data. These are all related to specific business scenarios. Data acquisition corresponding systems, such as trading systems, any core business operations of parents. Such as registration, payment, refund, transfer, shift and other core business, can be stable support. Even with system support, if there is a big deviation from the business, the data can not form an effective closed-loop system, and the collected data is not enough for analysis. Another example is the analysis of students' learning data. If students do not use electronic products to do questions and record data, follow-up work can not be carried out. And open a set of products, students do not like to use and do not need, also can not achieve data collection. Therefore, the key to data acquisition is whether there is a good business scenario and whether there is a reliable and stable system that fits the business scenario.

Data acquisition, but if the data information is separated from each other, it is difficult to integrate the analysis, thus affecting the scope of the impact, so the overall planning is very important. For example, as far as student data is concerned, no matter how many systems there are, for a certain system, student information is accessible. Only in this way can integrated analysis be possible.

What kind of data should the education industry collect? Taking the students as an example, the difficulty of the questions, the knowledge points of the questions, the time of the students' answers, the number of times of returning to the questions repeatedly, and the correct rate of answers are all indicators that need to be collected. When students watch videos, the progress of video playback, the number of times they look back and the length of time they play are all indicators that need to be collected.

2.2 Data storage
There is not much to be said about data storage. There is not much to be said about data storage. But anyone who has some experience in IT implementation knows that storage is nothing more than file storage and database storage. File storage is divided into stand-alone storage and distributed storage (such as HDFS), database storage is divided into relational database (such as MySQL, Oracle) storage and non-relational database (such as HBase) storage. For business scenarios, relational datasets are often used. In order to solve the problems of performance and concurrency, message queuing and master-slave synchronization are often introduced. In order to improve the efficient storage of data, MongoDB is often considered. In order to solve the problem of changing data, Redis is often considered.

For the analysis of data, the inevitable choice is to establish a data warehouse first. The establishment of data warehouse is always the same in every industry. The establishment of data warehouse should follow several principles: star architecture, theme-based design and hierarchical design. As far as layered design is concerned, it can be divided into ODS layer, DDS layer and EDS layer. ODS layer is also the original layer. Its data is the closest to the original data, and there is little difference between them. The main purpose is to isolate the ODS layer from the business system, so as to avoid affecting the performance of the business system because of the need of statistical analysis. Once the ODS layer is processed, it's time to design the DDS layer. What DDS needed to do was to separate different theme layers. What EDS layer needs to do is how to better visualize the display level. For data storage, consider using Greenplum or Hadoop.

2.3 Data processing

There are many methods of data processing, the best one is suitable for oneself. There is a big principle in data analysis, that is, it can not affect the normal operation of business. In the process of data processing, it is often necessary to write complex SQL scripts. If these scripts are not so well written, they may be fatal for business systems, such as causing resource contention, deadlock, etc., which are not desirable. For the storage mode of MySQL, isolation can be achieved through the master-slave synchronization mechanism of MySQL. For heterogeneous databases, it can be solved through Ali's open source project cannal. If we want to do real-time analysis, we can write the data to the message queue, and then calculate and analyze the data in real-time through spoutu and bolt computing unit of storm. If it is offline processing, there are many such tools, you can write a bunch of JAVA programs, such as DataStage, Information, etc., but more expensive, moving millions.

The essence of data processing is pre-processing. There are many scenarios in the field of analysis. In fact, there is no need for so much process data. The essence of pre-processing is to form the result data under the final dimension-index quantification system through certain analysis logic, thus helping to achieve rapid analysis and perspective.

2.4 Data analysis

The real value of data lies in analysis. If the refund rate in the transaction suddenly increases, we need to analyze what the root cause is, whether the system logic is not taken into account, or because of business errors. The monitoring and analysis of the conversion rate in channel delivery will determine the quality and cost of market delivery. Wal-Mart supermarkets have beer and diapers on similar shelves, affecting both sales. The analysis of market share will be helpful to the effective location of teaching and assistant institutions. The analysis of students'learning behavior will help improve their learning state and help them grow up healthily.

The core logic of data analysis lies in mining and discovering what from the data, and then putting forward a business proposition. With this proposition, the so-called data acquisition, storage and processing are all tools or methods for proposition realization.

SPSS, Clementine, Excel are commonly used tools for data analysis. Excel can help us analyze data quickly, such as PivotTable, Screening, Sorting, etc. If systematic data
calculation is carried out, R, Python and Spark frameworks are often chosen.

2.5 Data presentation

Data presentation is critical, and good data presentation can help analysts find problems quickly. For example, the trend chart of daily refund fee, when the curve fluctuates, we can clearly see what happened; scatter plot helps to analyze the correlation between the two indicators; radar plot helps to analyze where the weak points are; funnel plot helps to analyze sales transformation; mulberry plot helps to analyze the loss situation, etc.

Figures are important, tables are more pragmatic. The form is helpful for screening and comparative analysis. When presenting tables, we need to pay special attention to what indicators users are most concerned about, the selection relationship of user analysis, what users have seen, what levels they will analyze, and how the relationships among different levels are.

3 Reflection and summary

Big data can solve many unsolvable problems in the education system, but it is not omnipotent. On the data acquisition level, if children use electronic devices for too long, parents will worry about the impact on their children’s vision and whether the long-term use of electronic devices will affect their attention. Although parents know their children’s learning situation, will they handle it well? Is it possible to blame the child for hating the ironmaking, causing too much psychological pressure on the child, so that the child who was full of fun in learning lost interest in learning? Can teachers have a good grasp of their children’s ability to teach in accordance with their aptitude? These are all topics worth discussing.

In short, the application of big data in education industry is a double-edged sword, which will bring benefits and some problems. To solve these problems, we need to constantly optimize and improve. Therefore, big data is not achieved overnight, it takes time to demonstrate and evolve. Nevertheless, big data is still an important way to change the productivity of enterprises and improve the service chain. It is imperative to implement big data in education industry.

Reference


[Author’s Brief] Chen Wei, Associate Professor, Research Direction: Data Analysis and Mining. email: 474005517@qq.com

[Foundation Project] Project of Guangdong Provincial Higher Vocational Education Teacher Education Professional Teaching Steering Committee (No. 2016T078); School-level scientific research project of Guangdong Vocational College of Science and Technology (No. XJSC2016208)