Writing your thesis in LaTeX2e

Wangzhihui Mei

Supervisor: Zhifeng Wang

This thesis is submitted for the degree of

Master of Engineering

Central China Normal University Wollongong Joint Institute

Central China Normal University

January 2021

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgments.

Wangzhihui Mei
January 2021

Abstract

is paper implements a high school mathematics learning resource recommendation system based on knowledge tracking and factorization machine, which contains three core points. The first one is the design of data source, and proposes an automatic extraction method for test knowledge points. The whole process of high school mathematics knowledge mapping from corpus collection, construction, storage, and update is introduced as well as the automation process. The second point is the implementation of a GCN-based knowledge tracking algorithm, which takes into account the a priori structure of knowledge points as well as students' recent question records and knowledge mastery changes, and achieves a more excellent performance compared to other knowledge tracking models. The third point is based on a deep factorization machine, which solves the problem of sparsity of training data, and it takes the output of the knowledge tracking model as input and considers various other feature inputs to achieve learning resource recommendation.

Keywords: Learning Resource Recommendation System, Knowledge Graph，Knowledge Tracing，Factorization Machine，T，h

摘 要

中文摘要...

关键词：keyword1，keyword2，keyword3，keyword4

Acknowledgments

Publications

List your publications here...
Please remove this page if you do not have any...

目录

插图

表格

Nomenclature

Subscripts

crit Critical state

Acronyms / Abbreviations

ALU Arithmetic Logic Unit

BEM Boundary Element Method

CD Contact Dynamics

CFD Computational Fluid Dynamics

CK Carman - Kozeny

DEM Discrete Element Method

DKT Draft Kiss Tumble

DNS Direct Numerical Simulation

EFG Element-Free Galerkin

FEM Finite Element Method

FLOP Floating Point Operations

FPU Floating Point Unit

FVM Finite Volume Method

GPU Graphics Processing Unit

LBM	Lattice Boltzmann Method
LES	Large Eddy Simulation
MPM	Material Point Method
MRT	Multi-Relaxation Time
PCI	Peripheral Component Interconnect
PFEM	Particle Finite Element Method
PIC	Particle-in-cell
PPC	Particles per cell
RVE	Representative Elemental Volume
SH	Savage Hutter
SM	Streaming Multiprocessors
USF	Update Stress First
USL	Update Stress Last

Chapter 1

Introduction

1.1 Research Background and Significance

Artificial intelligence industry has developed rapidly in recent years, it is being commercialized in all aspects, triggering profound changes in various industries, and the future development of artificial intelligence will be the combination of key technologies and industries.[3] At present, AI technology has been implemented in many fields such as finance, medical and security, and the application scenarios are becoming more and more abundant. The commercialization of AI has played a positive role in accelerating the digitization of enterprises, improving the structure of the industrial chain, and increasing the efficiency of information utilization. The traditional education industry also tries to use AI technology to help the development of the industry. Every development of AI is accompanied by breakthroughs in research methods, and deep learning is one of the important representatives of the breakthroughs in machine learning technology in recent years. With the continuous extension of human AI research and application fields, AI will usher in more kinds of technology combination applications in the future. Artificial intelligence has also begun to be applied to the education industry, and the concept of intelligent education has emerged. Among the types of applications of AI technology in education, AI adaptive learning is the most widely used in all aspects of learning. In addition, due to China's large population base, the shortage of educational resources, the importance attached to education and other favorable factors intelligent adaptive learning system is expected to come later.

In recent years, domestic adaptive learning has begun to enter the minds of many people involved in education training and education investment. There are more and more education technology companies in the market that focus on adaptive learning tools. At the same time, many education companies have started to use adaptive learning as the main core function or main selling point of their products. The biggest advantage of adaptive education is that it can locate the knowledge gaps of each student. The adaptive learning platform will guide the student to the next most suitable learning content and activities for him. When students encounter a course that is too difficult or too low in the learning process, they can automatically adjust the difficulty of the course. Teachers can also analyze the knowledge gaps of each student based on the learning status evaluation report provided by the system, adjust the learning progress in real time, and provide personalized teaching for each student. So theoretically, adaptive learning is one of the potentially feasible solutions to the problem of "teaching to students according to their abilities" in online education. To make a practical adaptive learning system, I plan to use knowledge tracing to track students' learning status and use the factorization machine algorithm to calculate the relevance of topics to students to build such a test recommendation system.The current personalized learning resource recommendation system is one way of implementing adaptive learning, which is the subject of this paper.

In this paper, the study focuses on the recommendation of learning resources for the subject of high school mathematics. In this system, there are two aspects in general: on the one hand, scientific and targeted acquisition and tracing of students' knowledge state, and on the other hand, recommendation of personalized learning resources based on students' knowledge mastery state. We use the knowledge tracing algorithm of graph neural network to acquire and track students' knowledge states, and the factorization agent to try to combine the output of graph neural network with prior knowledge for resource recommendation.

1.2 Research Status

The dominant content of the research is knowledge tracing and recommendation system. Some advanced graph neural network algorithm is applied to finish the task. There have been some research advances and related applications in the area of knowledge tracing and factorization. We surveyed some existing knowledge tracing algorithms and applications, and some applications of factorization machine.

1.2.1 Property of high school Math

Disciplines and knowledge are closely related to each other, so that disciplinary knowledge denotes the specific knowledge contained in a particular field of study. Disciplines are referred to in this study only for specific subjects in the field of education, such as mathematics, language, chemistry and so on. The first step is to learn how to make the best use of the knowledge that is available. The knowledge is obtained from practice, so after learning it, it can also be applied to social practice. Scientific knowledge is declarative because it can be expressed in a series of symbols, words and diagrams; it is also procedural because it can be arranged and learned according to a specific logical order in the process of concrete learning.

Mathematics is a science specializing in the study of the relationship between quantities and spatial forms, its symbolic system is more complete, the formula structure is clear and unique, text and images and other expressions of language is also more vivid and intuitive.

The knowledge that learners need to learn mostly comes from the summaries of the experiences of their predecessors in practical activities. The learning process is a process of cognitive learning of the summarized knowledge and continuous digestion, adjustment and reorganization of the knowledge structure, so as to build a more perfect and suitable knowledge structure, as well as a process of integration with innovative thinking. Thus a good cognitive structure can promote the formation of knowledge structure, and a good knowledge structure can enrich the organization form of cognitive structure. Since the disciplinary knowledge structure consists of two parts: knowledge composition and knowledge dependency, we will analyze the disciplinary knowledge structure from these two aspects, knowledge structure and composition.

The knowledge that learners need to learn mostly comes from the summaries of the experiences of their predecessors in practical activities. The learning process is a process of cognitive learning of the summarized knowledge and continuous digestion, adjustment and reorganization of the knowledge structure, so as to build a more perfect and suitable knowledge structure, as well as a process of integration with innovative thinking. So a good cognitive structure can promote the formation of knowledge structure, and a good knowledge structure can enrich the organization form of cognitive structure. Since the disciplinary knowledge structure consists of two parts: knowledge composition and knowledge dependency, we will analyze the disciplinary knowledge structure from these two aspects.

Knowledge composition refers to the organization of knowledge within a subject area, which mainly includes knowledge points, knowledge blocks and knowledge systems.

- knowledge point: A point of knowledge is the smallest constituent unit of the knowledge structure of a discipline and is used to represent specific concepts.
- knowledge block: A knowledge block is a collection of one or more sets of knowledge points, also known as knowledge modules, in which knowledge blocks and knowledge blocks can be combined to form new knowledge modules, and a subset of knowledge blocks is called a knowledge sub-module.
- knowledge body: a body of knowledge is a structured system that is a combination of all the pieces of knowledge in a particular subject area.

Mathematics is a science that specializes in the relationship between quantity and spatial form. Its symbol system is more complete, the formula structure is clear and unique, and the language of expression such as words and images is more vivid and intuitive. The knowledge structure of senior secondary mathematics is a more logical and systematic knowledge system organized on the basis of the knowledge structure of junior secondary mathematics. This is because learning for any discipline needs to be based on the existing cognitive structure in order to progressively effective learning and skills training, so in the process of learning high school mathematics, you need to have a solid foundation of junior high school mathematics discipline knowledge. In the past few years, there have been a number of cases in which the students have been able to learn from each other.

- Highly abstract: Mathematics has a high degree of abstraction, because the discipline's knowledge system is built using many abstract knowledge concepts, and with the help of these concepts and knowledge to learn and expand thinking, forming new abstract conceptual knowledge. The abstraction of mathematics is reflected in the object is not concerned with the introduction of specific content, only the number of relationships between the spatial form. Therefore, abstraction in mathematics is different from abstraction in other disciplines in terms of both object and degree. There are also some differences between mathematics and the natural sciences, because in mathematics the accuracy of calculations, proofs, and inferences can only be verified using rigorous logical methods and cannot be tested by repetitive experiments, whereas in the natural sciences the verification is the opposite.

- Strict logic: The discipline of mathematics is very logical because any conclusion reached in mathematics requires rigorous logical reasoning and rigorous proof in order to be considered reasonable. However, mathematics is not the only discipline that possesses rigorous logic; other natural science studies of reasoning and proof must also possess a certain degree of logic. In mathematics, not all conclusions reached after reasoning and proof can be applied in practice, because many mathematical models are developed and mathematical conclusions drawn under ideal circumstances.

- Broad applicability: Mathematics is an important means and tool for us to participate in practical social activities or scientific research, and the study of mathematics is indispensable in all walks of life and in all areas of society. Therefore, mathematics has a wide range of applications and has become an important basis for the development of modern science.

1.2.2 Knowledge relation

Knowledge relations represent the connections between knowledge points (or between knowledge blocks and knowledge chunks) in the discipline knowledge structure. It is through these connections that different knowledge points can be formed into knowledge blocks, and different knowledge blocks can be combined to form the whole disciplinary network knowledge structure system. There are many different kinds of knowledge relationships, so that different definitions of knowledge relationships lead to different knowledge structures. Therefore, in order to unify the definition of knowledge relations, we divided them into general relations and special relations based on the general and special characteristics of discipline knowledge structure. The special relationships represent the unique knowledge relationships of a particular discipline, while the universal relationships represent the general relationships of any discipline. Secondly, according to the demands of knowledge graphping research, we divide universal relations into six kinds of knowledge relations: synonymous, fraternal, antecedent, consequent, inclusive and antagonistic; and special relations into four kinds of knowledge relations: detailed, transformative, causal and correlative.

- tautology: Expresses the relationship between two points of knowledge that have the same meaning as what is being described, e.g. regular and equilateral triangles.

- fraternity: Expresses the relationship between two knowledge points that have the same parent class.

- predecessor: It means that you need to finish learning knowledge point A before learning knowledge point B, that is, $A \rightarrow B$ is a precursor relationship.

- successor: denotes the inverse of the antecedent relationship, i.e., $B \rightarrow A$ is the successor relationship.

- containment: Indicates that knowledge point B is included in the definition of knowledge point A, i.e., $A \rightarrow B$ is an inclusion relationship.

- antagonism: From a certain point of view, knowledge point A is incompatible with knowledge point B, i.e. $A \leftrightarrow B$ is an antagonistic relationship.

- refinement: A grammatical analysis of the definition of knowledge point A leads to knowledge point B, where $A \leftrightarrow B$ is a detailed relationship

- transformation: denotes that knowledge point A and knowledge point B can be transformed to each other under certain conditions, i.e., $A \leftrightarrow B$ is a transformation relationship.

- causation: denotes that knowledge point A can be deduced from knowledge point B as a known condition, i.e., $A \leftrightarrow B$ is a causal relationship.

- relation: Indicates that there is a relationship between the definitions of Knowledge Point A and Knowledge Point B, but the relationship is not explicitly specified, i.e., $A \leftrightarrow B$ are correlated.

In the process of constructing the discipline knowledge structure, firstly, we need to analyze the current discipline knowledge content, teaching objectives, teaching objects, teaching strategies and discipline characteristics in detail; secondly, we divide the whole discipline knowledge system into several knowledge modules, and then we divide each knowledge module into several knowledge points; finally, with reference to the above ten kinds of knowledge relationships and the knowledge relationships extracted from data sources, we can determine and establish the relationships between knowledge modules and knowledge modules, between knowledge modules and knowledge points, and between knowledge points and knowledge points, so as to form a complete discipline knowledge system structure.

1.2.3 Knowledge tracing algorithms

Knowledge Tracing is a technique that models students' knowledge acquisition based on their past answers to obtain a representation of their current knowledge state. The task is to automatically track the change of students' knowledge level over time based on their historical learning trajectory, in order to be able to accurately predict the students' performance in future learning and to provide appropriate learning tutoring. In this process, the knowledge space is used to describe the level of student knowledge acquisition. A knowledge space is a collection of concepts, and a student's mastery of a part of a collection of concepts constitutes the student's mastery of knowledge. Some educational researchers argue that students' mastery of a particular set of related knowledge points will affect their performance on the exercise, i.e., the set of knowledge that students have mastered is closely related to their external performance on the exercise.

The task of knowledge tracing is to model the student's knowledge mastery state based on the student's answer record, which is usually a time series, and in some business scenarios is time-independent, so that we can accurately predict their future answers and make reference for future intelligent questioning based on this to avoid giving students too difficult or too easy questions. Specifically, suppose a student's answer record is $x_0, x_1, ..., x_t$, and we are going to predict the next interaction x_{t+1}, usually one interaction $x_t = (q_t, a_t)$, q_t represents the right or wrong situation of that student's answer to the question a_t.

There are several kinds of knowledge tracing algorithms:

- Bayesian knowledge tracing(BKT): Bayesian knowledge tracing is an early and commonly used knowledge tracing model, BKT uses user interaction modeling with real-time feedback to model a learner's potential knowledge state as a set of binary variables, each representing whether or not a knowledge point is understood, and there are dynamic changes in mastery of knowledge points as students continue to practice, BKT maintains binary variables of knowledge point proficiency by using Hidden Markov Models (HMM), the original BKT model does not take into account students' knowledge forgetting, and related studies address students' guesses, personal vivid knowledge mastery and problem difficulty factors on BKT[10].

- Deep Knowledge Tracing(DKT): The DKT model applies neural networks to the knowledge tracing task for the first time[6], using an LSTM model to track the dynamics of student knowledge proficiency over time, and to learn the potential vector representation of student knowledge proficiency directly from the data. The advantage of DKT is that it can record knowledge over a longer period of time based on students' recent answers. In addition, it can update the knowledge state based on each answer, only the last implicit state needs to be saved, no double counting is required, and it is suitable for online deployment. It does not require domain knowledge, works with any user answer dataset and automatically captures associations between similar questions. The disadvantage of DKT is that the model output fluctuates greatly when the answer sequence is disrupted, i.e., the same questions and the same responses yield different knowledge states when the answer sequence is inconsistent. Due to the above-mentioned problems and the fact that students do not necessarily have continuous consistency in their k nowledge during the answer process, it leads to bias in the prediction of students' knowledge states influenced by the sequence. There is also the black box problem, which sometimes leads to the strange situation that the first correct answer leads to a high prediction probability for all subsequent ones, while the first wrong answer leads to a low prediction probability for all subsequent ones.

- Dynamic Key-Value Memory Networks for Knowledge Tracing(DKVMN): Dynamic Key-Value Memory Networks for Knowledge Tracing (DKVMN) was proposed in 2017 by Jian Jian of the Chinese University of Hong Kong[11]. Based on the strengths and weaknesses of BKT and DKT and using the memory augmentation neural network approach, the Dynamic Key-Value Memory Networks (DKVMN) is proposed. It borrows ideas from memory-enhanced neural networks and combines the advantages of BKT and DKT. DKVMN stores all knowledge points with a static matrix key and a dynamic matrix value to store and update the student's knowledge state. In the DKVMN paper, they compare DKVMN with DKT and a sophisticated version of BKT, BKT+. They found that DKVMN achieves excellent performance and is the most advanced model in the KT domain. In addition to improved performance, it has several other advantages over LSTM, including prevention of overfitting, a smaller number of parameters, and automatic discovery of similar practice questions by underlying concepts. In addition, Chaudhry R[2] improves the performance of DKVMN by jointly training request cue prediction with knowledge tracing through multi-task learning.

1.2.4 Factorization Machine

The factorization machine model is a factorization model that can be used in large scale sparse data scenarios[8]. The solution of this model is linear in time complexity and he can solve it directly using raw data without relying on support vectors like SVM. In addition, FM is a general model that can be used on any real data and can do tasks such as classification and regression and even sorting. The idea is that the idea is to add a linear combination of two features to linear regression, and the way to solve the linear combination is to use a matrix decomposition. FFM is an improvement on FM by adding the concept of Field[5], that is, the class to which each feature belongs. Suppose there are f Fields. Then each feature has to have f hidden vectors. When two features are crossed, the dot product of each feature and the vector corresponding to the other Field is used. In this way, it is ensured that the same Field does the same thing for the same feature, and the features of different Fields do different things for the same feature. In addition, there is also a deep learning version of the factorization machine algorithm[4], and this model is improved based on wide and deep. First, the model includes FM and DNN parts, which is a parallel structure, and FM and DNN share the same input (embedding). The mapping vector from the field to the embedding layer is exactly the vector learned by the FM layer. It has the advantage that it does not require pre-training and can learn the intersection of low and high dimensional features.

1.3 Research Objectives and Content

The purpose of this study is to build a high school mathematics learning resource recommendation system based on knowledge tracing and factorization machine algorithm. We use knowledge tracing to model students' knowledge states, which outputs a graphical knowledge state vector, which we use as the next-level input, considering students' individualized differences and knowledge forgetting process, and apply the factorization machine algorithm to the resource recommendation system. For knowledge tracing, we build a graph neural network-based knowledge tracing model, which can well characterize the intrinsic connections of knowledge points in mathematics subjects considering that the knowledge points are a graph-like structure, and output a graph knowledge vector matrix, which can also effectively characterize the connections between problems and knowledge points. The output of the knowledge tracing model is then passed through a factorization machine algorithm to obtain the recommendation degree of the learning resources and output a vector of recommendation weights for different learning resources.

1.4 Thesis Organization and Structure

Chapter 1 of this paper is an introduction. It introduces the research background of the study, current industry-related research progress and the focus of the study. Then it leads to the three core points of this paper: learning resource representation, knowledge tracing and resource recommendation.

Chapter 2 of this paper concentrates on learning resource representation, which addresses storing learning resources through knowledge graphs. This paper explores some concepts of knowledge graphs, related studies, and then gives the process of knowledge graph building. It is also demonstrated that knowledge graph building can effectively characterize the a priori intrinsic features of subject knowledge.

Chapter 3 of this paper gives a knowledge tracing model of pre-trained graph neural network, which better characterizes the graph-like properties of knowledge. It is able to transform the knowledge tracing task into a time-series node-level classification problem in GNNs. Since the knowledge graph structure is not explicitly provided in most cases, we present various implementations of the graph structure. Empirical tests on two open datasets show that the method improves the prediction of student performance without any additional information and shows more interpretable predictions. The inclusion of pre-training is also attempted during the experiments, which can greatly improve the training efficiency and performance.

Chapter 4 of this paper proposes the application of a deep factorization machine algorithm with knowledge tracing model data as input to build a learning resource recommendation system. The output is a weight vector of top n, characterizing the recommended resources of top n.

Chapter 5 of this paper presents the conclusion.

Chapter 2

Learning Resource Database based on Knowledge Graph

2.1 Research Motivation

As the starting point of a recommendation system, the data source is often the first step to consider. 在这里我们

A knowledge graph is a map that serves as a kind of knowledge domain mapping, which shows the knowledge development process and structure. Visualization techniques are used to describe knowledge resources and their carriers, to mine, analyze, construct, map and display knowledge and their interconnections. Also knowledge graphs, as an important part of artificial intelligence, are now playing an increasingly important role in recommendation systems and question and answer systems.

In this section, this paper uses the knowledge graph to store the knowledge points and questions of high school mathematics as the data source for the recommendation system. It serves as a supplementary input to the knowledge tracing model section and provides a priori knowledge of the subject knowledge points.

Firstly, the basic principles and common techniques of knowledge graph are introduced, and then a knowledge graph construction method for automatic extraction of knowledge points of test questions is proposed, and the knowledge graph construction and knowledge point extraction, corpus acquisition and preprocessing, knowledge point identification and knowledge graph updating techniques are introduced.

In this paper, we will use the knowledge graph in constructing the knowledge graph of mathematics domain and building the student user model. In the construction of the knowledge graph of high school mathematics domain, the basic concepts and keywords involved in the textbooks are abstracted into knowledge point ontologies, the knowledge point keywords in high school mathematics exercises are identified by the named entity recognition method, and then the abstracted knowledge points are associated with the keywords to build the knowledge graph of high school mathematics. A large number of knowledge points in high school mathematics are associated with the degree of mastery of each knowledge point by individual students, which is built into a student user model, and the abstraction of the student user model into a knowledge graph will facilitate the personalized recommendation of exercises.

2.2 Research Status

2.3 Basic theory of knowledge graph

The Knowledge Graph is a new concept introduced by Google in 2012. Knowledge graph is essentially a knowledge base of Semantic Network, which describes conceptual entities and their relationships in the objective world in a structured form. From the beginning of oogle search, to nowadays chatbots, big data risk control, securities investment, intelligent medical care, adaptive education, recommendation system, all use knowledge graph.

2.3.1 Representation

Knowledge graphs focus on concepts, entities and their relationships, where entities are things in the objective world and concepts are generalizations and abstractions of things with the same properties. Ontology is the basis of knowledge representation of knowledge graph, which can be formally represented as $O = \{C, H, P, A, I\}$, where C is the set of concepts, such as transactional concepts and event-like concepts, H is the set of contextual relations of concepts, P is the set of attributes, which describes the features possessed by concepts, A is the set of rules, which describes the domain rules, and I is the set of instances, which describes the instance-attribute-value.

The common knowledge graph representations are Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), Web Ontology Language (OWL), etc.

1. resource description framework RDF is the most commonly used symbolic semantic representation model, which provides a unified standard for describing entities/resources. the basic model of RDF is a directed labeled graph, each edge on the way corresponds to a subject-predicate object triad, and a triad corresponds to a statement of an event. the RDF consists of nodes and edges, the nodes represent entities/resources, attributes, and the edges represent entities and the relationship between entities and attributes. Figure 2.1 shows the relationship between entities and attributes.

2. lightweight schema language RDFS is an extension of RDF by adding the definition of class properties and other Schema layers on top of the objective events provided by RDF. rdfs is mainly used to define term sets, class sets and property sets, mainly including classes, subclasses, properties, subclass properties, domains, scopes and other primitives, which can build the basic class hierarchy and These clauses can build the basic class hierarchy and attribute system.

3. OWL is the core of the Semantic Web technology stack, which provides fast and flexible data modeling capability and efficient automatic reasoning capability.

Fig. 2.1 Structure of Triad

2.3.2 Storage

There are mainly two kinds of storage for knowledge graphs, one is based on the academic idea of RDF, which is based on triples car storage: the other is based on the industrial idea of attribute graph, which is based on attribute graph storage.

Using RDF for storage, mainly RDF serialization methods, such as RDF/XML, N-Triples, Turtle, RDFa, JSON-LD, etc.

The most important feature of the graph database for storing knowledge graphs is that the graph database can store not only the nodes and relations of triples, but also the attributes of their nodes and relations together, and the graph can be traversed efficiently, which is convenient for publishing data and retrieval of knowledge graphs.

2.4 The construction of knowledge graph

In order to convert a large number of standard questions into feature questions in a timely manner, this paper proposes an automatic knowledge extraction method based on knowledge graph for test questions. The method uses natural language processing techniques such as named entity recognition, keyword extraction and calculation of semantic similarity of keywords to collect knowledge points and mathematics-related vocabulary in high school mathematics domain and build up a knowledge map of high school mathematics. The mathematics test questions are divided into words and entities by customized dictionaries and rules, and the candidate keywords are extracted and brought into the knowledge map of high school mathematics for querying.

2.4.1 The extraction of knowledge point

In this paper, the architecture for building the knowledge map of high school mathematics is mainly divided into four modules: corpus collection module, pre-processing module, knowledge map building module and knowledge extraction system module. The high school mathematics mapping consists of two entities: mathematical knowledge points and mathematical vocabulary related to mathematical knowledge points, which are mainly used to provide automatic knowledge extraction function for the construction of intelligent question bank later. The question bank not only contains a large number of mathematical knowledge points, but also contains a lot of mathematical vocabulary related to mathematical knowledge points. The mathematical vocabulary does not only stop at the mathematical concept ontology, mathematical textbook chapter names, mathematical knowledge points, etc., but also includes words or vocabulary that can deduce the knowledge points of the test questions. For example, if the operation symbol of " $\cap$ " appears in the question, it can be deduced that the question is probably about "set", "set operation", "set intersection operation", "set element interdependence", "set interval", "set interval", "set interval", "set interdependence", "set interval", etc. "The pre-processing module is to standardize the imported data corpus to form structured data, and then to identify the named entities, which mainly involves the process of entity identification and relationship extraction. High school mathematics is relatively a semi-closed and semi-open subject, the mathematical knowledge is relatively closed, but the mathematical test questions are very different, but the final solution still comes back to the test mathematical knowledge, so it is semi-open. The relevant mathematical knowledge points and concepts obtained through knowledge extraction need to be evaluated by experts to create a valuable and credible knowledge map. The knowledge extraction system plays a key role in the construction of the intelligent question bank by tagging the questions obtained by crawlers on the web after processing.

2.4.2 Corpus data preprocessing

After importing the textbook and web crawler corpus, although a large amount of valuable text, image information and structured data such as question stems and parses in the question bank have been obtained, there are a large number of mathematical formulas in these data, and these mathematical formulas are not standardized because of the platform, and there is no uniform storage format, so it is necessary to standardize these text data containing mathematical formulas.

- standardization: The standardization process is mainly for the pictures and mathematical formulas in the corpus, many crawlers get the test information through the pictures, the test information stored in the pictures is very unfavorable to the word separation and naming entity recognition, and the mathematical formulas also exist in a variety of storage methods, currently more common storage and display of mathematical formulas are Office comes with the formula editor plug-in, Mathtype formula editor, Mathml, Latex and so on. For picture information, we use OCR technology for text recognition, and for mathematical formulas, we use Latex format. Mathpix is a software for quickly recognizing mathematical formulas stored in pictures and converting them into Latex format with high recognition accuracy and efficiency, and Latex can extract some useful semantic information for later knowledge extraction when expressing mathematical formulas.

- User dictionaries and stopwords: User dictionaries, also known as user-defined dictionaries, are mainly used to enhance the disambiguation and error correction ability of the subscripts by manually adding subscripts rules in the process of named entity recognition, and the subscripts recognition will give priority to the words in the user dictionaries for subscripts after adding the user dictionaries. At present, there is no mature lexical corpus in the field of mathematics, so it is important to build a user lexicon related to mathematics. For example, if we do not add a user lexicon for "function analytic", we will get [function, analytic, equation] if we use the popular third-party corpus Jieba, but we do not want to see such a result when we add the field "analytic" to the user lexicon. When the "parser" field is added to the user's dictionary, the result will be [function,parser]. Deactivated words are also called "dummy words in computer search, non-search words". In search engines, in order to save space and search efficiency, certain words or phrases are usually automatically ignored in search requests, and these words or phrases are called deactivated words. The deactivation dictionary is a filter composed of a number of deactivation words, in the named entity recognition of the word, the system can be based on the deactivation dictionary to filter out some of the words or words that are not useful, to improve the accuracy of the word and the system computing efficiency. Deactivated words mainly include common pronouns, inflectional auxiliaries, adverbs, prepositions, conjunctions, etc., which usually have no obvious meaning of their own and only have a certain role when they are put into a complete sentence, such as: [you, I, he, this, that, the, in, then], etc.

- Tokenization: In the process of constructing mathematical knowledge graphs, word separation is mainly used to obtain mathematical knowledge points and related mathematical vocabulary, so there is no strict requirement on the lexicality of the words obtained by word separation. In this paper, we use Hanlp natural language processing toolkit to classify the text by perceptual machine. Hanlp has the features of perfect function, high performance, clear architecture, new corpus and customizable. and can recognize new words

2.5 Experiment

2.5.1 Dataset and Environment

The accuracy of the knowledge point extraction directly affects the construction of the feature database. The experiment selected 100,000 and 500 high school mathematics test questions with knowledge points in the web crawler as samples, and divided them into two parts for the automatic extraction of knowledge points without and with expert audit. As shown in Table , the accuracy of knowledge point extraction is 72.3% when there is no expert review and 96.6% when there is expert review. This experiment shows that the real-time updating of the knowledge map of high school mathematics in the process of automatic knowledge point extraction is the key to improve the accuracy of knowledge point extraction.

	Without expert review	With expert review
Sample Size	100000	500
Accuracy	72.3%	96.6%

表 2.1 The accuracy comparison table

2.5.2 Performance Indicators

2.5.3 Design of Experiment

2.5.4 Experiment Review

2.5.5 The Exercise Embedding

2.6 Summary

This chapter introduces the automatic knowledge extraction method of test questions based on knowledge graph. The intelligent algorithm based on individual features proposes to convert the standard question bank into a feature bank with knowledge feature representation, and defines each dimension in the feature bank, constructs the individual feature bank for users according to individual knowledge points, and explains the recommendation algorithm of test questions in the intelligent question bank; the automatic extraction method of knowledge points of test questions based on knowledge map introduces the whole process of knowledge map of high school mathematics from corpus collection, construction, storage and update. This chapter also introduces the automatic extraction process of knowledge points of high school mathematics test questions. This chapter also presents a comparative experiment on the objectivity of feature database construction and the effectiveness of automatic knowledge point extraction, and the related analysis of the experimental results.

Chapter 3

Knowledge Tracing Model Based on Graph Attention Networks

3.1 Motivation

This section is a core part of this recommendation system, which is to obtain the student's knowledge mastery status by means of knowledge tracking. Knowledge tracking is to model students' knowledge mastery based on their past answer records to obtain students' knowledge status. Models of knowledge tracking are abundant, and early models of knowledge tracking are generally based on Bayesian knowledge tracking (BKT) with first-order Markov models, which are based on the assumption that students' knowledge states are represented by a set of binary variables and have also achieved good results. In 2015, Piech et al.[6] proposed a deep knowledge tracking model (DKT), which for the first time applied recurrent neural networks to a knowledge tracking task for tracking the dynamic change of students' knowledge proficiency over time and learning students' potential vector representation for knowledge proficiency, marking the prologue of knowledge tracking research based on neural network models. Later, there was the improved model Exercise-Enhanced Recurrent Neural Network (EERNN)[9] for mining the textual information of question stems for the model cold-start problem, which also utilized the attention mechanism to consider the problem of topic similarity. Later Zhang et al. proposed the Dynamic Key-Value Memory Networks model[11], which considers the relationship between knowledge concepts and skills, was proposed, which builds a Key-Value model for the skill-question binary relationship and achieves better results. It also has good interpretability, which intuitively reflects the logical relationship between skill mastery and question answering, and is in line with the relevant educational psychology principles. For these models there are also some studies that consider the influence of some other factors, such as the introduction of forgetting problem, multiple knowledge point problem and a priori skill mapping considerations. However, the existing models often do not consider enough the connection between knowledge points, either simply as mutually independent nodes or as simple hierarchical models, but in fact, knowledge points are graph-like structures, in which case, using graph neural networks to characterize the relationship between knowledge points and training their embedding is a better solution.

3.2 Proposed Model

3.2.1 Algorithm Overview

3.2.2 Graph Neural Networks

3.2.3

3.3 Second Section of the Third Chapter

3.4 The Layout of Formal Tables

This section has been modified from "Publication quality tables in LaTeX*" by Simon Fear.

The layout of a table has been established over centuries of experience and should only be altered in extraordinary circumstances.

When formatting a table, remember two simple guidelines at all times:

1. Never, ever use vertical rules (lines).

2. Never use double rules.

These guidelines may seem extreme but I have never found a good argument in favour of breaking them. For example, if you feel that the information in the left half of a table is so different from that on the right that it needs to be separated by a vertical line, then you should use two tables instead. Not everyone follows the second guideline:

表 3.1 A badly formatted table

	Species I		Species II	
Dental measurement	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

表 3.2 A nice looking table

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

There are three further guidelines worth mentioning here as they are generally not known outside the circle of professional typesetters and subeditors:

3. Put the units in the column heading (not in the body of the table).

4. Always precede a decimal point by a digit; thus 0.1 *not* just .1.

5. Do not use 'ditto' signs or any other such convention to repeat a previous value. In many circumstances a blank will serve just as well. If it won't, then repeat the value.

A frequently seen mistake is to use '\begin{center}' … '\end{center}' inside a figure or table environment. This center environment can cause additional vertical space. If you want to avoid that just use '\centering'

表 3.3 Even better looking table using booktabs

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

Chapter 4

My Third Chapter

4.1 First Section of the Third Chapter

And now I begin my third chapter here ...

And now to cite some more people Read [7], Ancey et al. [1]

4.1.1 First Subsection in the First Section

... and some more

4.1.2 Second Subsection in the First Section

... and some more ...

First Subsub Section in the Second Subsection

... and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it ...

4.1.3 Third Subsection in the First Section

... and some more ...

First Subsub Section in the Third Subsection

...and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it and some more and some more and some more and some more and some more and some more and some more ...

Second Subsub Section in the Third Subsection

...and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it ...

4.2 Second Section of the Third Chapter

and here I write more ...

4.3 The Layout of Formal Tables

This section has been modified from "Publication quality tables in LaTeX*" by Simon Fear.

The layout of a table has been established over centuries of experience and should only be altered in extraordinary circumstances.

When formatting a table, remember two simple guidelines at all times:

1. Never, ever use vertical rules (lines).

2. Never use double rules.

These guidelines may seem extreme but I have never found a good argument in favour of breaking them. For example, if you feel that the information in the left half of a table is so different from that on the right that it needs to be separated by a vertical line, then you should use two tables instead. Not everyone follows the second guideline:

There are three further guidelines worth mentioning here as they are generally not known outside the circle of professional typesetters and subeditors:

3. Put the units in the column heading (not in the body of the table).

4. Always precede a decimal point by a digit; thus 0.1 *not* just .1.

表 4.1 A badly formatted table

	Species I		Species II	
Dental measurement	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

表 4.2 A nice looking table

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

5. Do not use ‘ditto’ signs or any other such convention to repeat a previous value. In many circumstances a blank will serve just as well. If it won’t, then repeat the value.

A frequently seen mistake is to use ‘\begin{center}’ … ‘\end{center}’ inside a figure or table environment. This center environment can cause additional vertical space. If you want to avoid that just use ‘\centering’

表 4.3 Even better looking table using booktabs

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

Chapter 5

My Third Chapter

5.1 First Section of the Third Chapter

And now I begin my third chapter here ...

And now to cite some more people Read [7], Ancey et al. [1]

5.1.1 First Subsection in the First Section

... and some more

5.1.2 Second Subsection in the First Section

... and some more ...

First Subsub Section in the Second Subsection

... and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it ...

5.1.3 Third Subsection in the First Section

... and some more ...

First Subsub Section in the Third Subsection

. . . and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it and some more and some more and some more and some more and some more and some more and some more . . .

Second Subsub Section in the Third Subsection

. . . and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it . . .

5.2 Second Section of the Third Chapter

and here I write more . . .

5.3 The Layout of Formal Tables

This section has been modified from "Publication quality tables in LaTeX*" by Simon Fear.

The layout of a table has been established over centuries of experience and should only be altered in extraordinary circumstances.

When formatting a table, remember two simple guidelines at all times:

1. Never, ever use vertical rules (lines).

2. Never use double rules.

These guidelines may seem extreme but I have never found a good argument in favour of breaking them. For example, if you feel that the information in the left half of a table is so different from that on the right that it needs to be separated by a vertical line, then you should use two tables instead. Not everyone follows the second guideline:

There are three further guidelines worth mentioning here as they are generally not known outside the circle of professional typesetters and subeditors:

3. Put the units in the column heading (not in the body of the table).

4. Always precede a decimal point by a digit; thus 0.1 *not* just .1.

表 5.1 A badly formatted table

	Species I		Species II	
Dental measurement	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

表 5.2 A nice looking table

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

5. Do not use 'ditto' signs or any other such convention to repeat a previous value. In many circumstances a blank will serve just as well. If it won't, then repeat the value.

A frequently seen mistake is to use '\begin{center}' … '\end{center}' inside a figure or table environment. This center environment can cause additional vertical space. If you want to avoid that just use '\centering'

表 5.3 Even better looking table using booktabs

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

Chapter 6

Conclusion and Future Work

In this paper, a test question recommendation system based on graph neural network combined with knowledge mapping and factorization algorithm is established, which generally accomplishes the design goals and is original in the three stages of the recommendation system-data source generation, knowledge tracking and knowledge recommendation, etc. This paper adopts the current more popular graph neural network model and makes some improvements to the existing model to solve the tasks of knowledge analysis of test questions and knowledge state tracking of students, and achieves the desired results in the experiments with some performance improvements to the existing model.

Among them, in the data source part, they are input to the initial database by crawlers and manual input, and then the automatic knowledge point analysis model of test questions based on text mining and graph neural network iterative learning is used to complete the knowledge point label mining of test questions, and they then establish the knowledge map of high school mathematics by some means of knowledge map construction, which solves the "zero resource " problem. In terms of the knowledge tracking of students, this paper innovatively applies the graph self-attentive network to this task, and through the improvement of the model and the consideration of several factors, the task has been improved somewhat compared with the existing model and has improved in performance. Finally, by labeling the knowledge of the test questions and tracking the students' knowledge, the factorization machine algorithm is used to complete the learning resource recommendation, which solves the cold start problem and accomplishes the design goal of adaptive learning.

In future work, the sequential nature of the model can be solved by combining other graph neural network models or adding some memory mechanisms.

References

[1] Ancey, C., Coussot, P., and Evesque, P. (1996). Examination of the possibility of a fluid-mechanics treatment of dense granular flows. *Mechanics of Cohesive-frictional Materials*, 1(4):385–403.

[2] Chaudhry, R., Singh, H., Dogga, P., and Saini, S. K. (2018). Modeling hint-taking behavior and knowledge state of students with multi-task learning. *International Educational Data Mining Society*.

[3] Cui, W., Xue, Z., and Thai, K.-P. (2018). Performance comparison of an ai-based adaptive learning system in china. In *2018 Chinese Automation Congress (CAC)*, pages 3170–3175. IEEE.

[4] Guo, H., Tang, R., Ye, Y., Li, Z., and He, X. (2017). Deepfm: a factorization-machine based neural network for ctr prediction. *arXiv preprint arXiv:1703.04247*.

[5] Juan, Y., Zhuang, Y., Chin, W.-S., and Lin, C.-J. (2016). Field-aware factorization machines for ctr prediction. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 43–50.

[6] Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., and Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in neural information processing systems*, 28:505–513.

[7] Read, C. J. (1985). A solution to the invariant subspace problem on the space l_1. *Bull. London Math. Soc.*, 17:305–317.

[8] Rendle, S. (2010). Factorization machines. In *2010 IEEE International Conference on Data Mining*, pages 995–1000. IEEE.

[9] Su, Y., Liu, Q., Liu, Q., Huang, Z., Yin, Y., Chen, E., Ding, C., Wei, S., and Hu, G. (2018). Exercise-enhanced sequential modeling for student performance prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

[10] Yudelson, M. V., Koedinger, K. R., and Gordon, G. J. (2013). Individualized bayesian knowledge tracing models. In *International conference on artificial intelligence in education*, pages 171–180. Springer.

[11] Zhang, J., Shi, X., King, I., and Yeung, D.-Y. (2017). Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web*, pages 765–774.

附录 A

How to Install LaTeX

Windows OS

TeXLive package - full version

1. Download the TeXLive ISO (2.2GB) from
   https://www.tug.org/texlive/

2. Download WinCDEmu (if you don't have a virtual drive) from
   http://wincdemu.sysprogs.org/download/

3. To install Windows CD Emulator follow the instructions at
   http://wincdemu.sysprogs.org/tutorials/install/

4. Right click the iso and mount it using the WinCDEmu as shown in
   http://wincdemu.sysprogs.org/tutorials/mount/

5. Open your virtual drive and run setup.pl

or

Basic MikTeX - TeX distribution

1. Download Basic-MiKTeX(32bit or 64bit) from
   http://miktex.org/download

2. Run the installer

3. To add a new package go to Start » All Programs » MikTex » Maintenance (Admin) and choose Package Manager

4. Select or search for packages to install

TexStudio - TeX editor

1. Download TexStudio from
   http://texstudio.sourceforge.net/#downloads

2. Run the installer

Mac OS X

MacTeX - TeX distribution

1. Download the file from
   https://www.tug.org/mactex/

2. Extract and double click to run the installer. It does the entire configuration, sit back and relax.

TexStudio - TeX editor

1. Download TexStudio from
   http://texstudio.sourceforge.net/#downloads

2. Extract and Start

Unix/Linux

TeXLive - TeX distribution

Getting the distribution:

1. TexLive can be downloaded from
   http://www.tug.org/texlive/acquire-netinstall.html.

2. TexLive is provided by most operating system you can use (rpm,apt-get or yum) to get TexLive distributions

Installation

1. Mount the ISO file in the mnt directory

```
mount -t iso9660 -o ro,loop,noauto /your/texlive####.iso /mnt
```

2. Install wget on your OS (use rpm, apt-get or yum install)

3. Run the installer script install-tl.

```
cd /your/download/directory
./install-tl
```

4. Enter command ‘i’ for installation

5. Post-Installation configuration:
http://www.tug.org/texlive/doc/texlive-en/texlive-en.html#x1-320003.4.1

6. Set the path for the directory of TexLive binaries in your .bashrc file

For 32bit OS

For Bourne-compatible shells such as bash, and using Intel x86 GNU/Linux and a default directory setup as an example, the file to edit might be

```
edit $~/.bashrc file and add following lines
PATH=/usr/local/texlive/2011/bin/i386-linux:$PATH;
export PATH
MANPATH=/usr/local/texlive/2011/texmf/doc/man:$MANPATH;
export MANPATH
INFOPATH=/usr/local/texlive/2011/texmf/doc/info:$INFOPATH;
export INFOPATH
```

For 64bit OS

```
edit $~/.bashrc file and add following lines
PATH=/usr/local/texlive/2011/bin/x86_64-linux:$PATH;
export PATH
MANPATH=/usr/local/texlive/2011/texmf/doc/man:$MANPATH;
export MANPATH
INFOPATH=/usr/local/texlive/2011/texmf/doc/info:$INFOPATH;
export INFOPATH
```

Fedora/RedHat/CentOS:

```
sudo yum install texlive
sudo yum install psutils
```

SUSE:

```
sudo zypper install texlive
```

Debian/Ubuntu:

```
sudo apt-get install texlive texlive-latex-extra
sudo apt-get install psutils
```

附录 B

Installing the CUED Class File

LaTeX.cls files can be accessed system-wide when they are placed in the <texmf>/tex/latex directory, where <texmf> is the root directory of the user’s TeXinstallation. On systems that have a local texmf tree (<texmflocal>), which may be named “texmf-local” or “localtexmf”, it may be advisable to install packages in <texmflocal>, rather than <texmf> as the contents of the former, unlike that of the latter, are preserved after the LaTeXsystem is reinstalled and/or upgraded.

It is recommended that the user create a subdirectory <texmf>/tex/latex/CUED for all CUED related LaTeXclass and package files. On some LaTeXsystems, the directory look-up tables will need to be refreshed after making additions or deletions to the system files. For TeXLive systems this is accomplished via executing “texhash” as root. MIKTeXusers can run “initexmf -u” to accomplish the same thing.

Users not willing or able to install the files system-wide can install them in their personal directories, but will then have to provide the path (full or relative) in addition to the filename when referring to them in LaTeX.