Information Processing and Experiment Platform Simulation on Social Network

Yan Yan^{*}, Faguo Zhou, Penghao Min and Jie Zhang

Department of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing 100083, China.

Abstract

With the widespread use of mobile devices, e-commerce websites and related services are becoming increasingly mature, such as buying clothes online, booking airline tickets, taking a taxi and buying groceries. Online shopping highly facilitate human beings in food, clothing, shelter and transportation. But at the same time, compared with the traditional way, due to the inability to try out the goods and the lack of intuitive perception, the product selection process of online shopping depends largely on the feedback of the purchased users, an important part of which is user reviews. Large-scale reviews are posted by a huge number of users based on the content and quality of the shopping experience goods, however, not all online reviews have reference value, and how to deal with this information is the main problem faced by big data. To address this problem, we build the labeled Chinese dataset of helpfulness analysis in the e-commerce domain. Meanwhile, we create an experimental platform based on the Jupyter Notebook tool. The platform uses the dataset constructed above as experimental data and as a case study for teaching helpfulness analysis. The platform integrates online learning module, interactive sharing module, and teaching management module, and has advantages such as online programming, resource integration, and interactive experience. In addition to the basic theoretical knowledge, the platform has also designed a series of natural language processing cases from the shallower to the deeper, to cultivate students' comprehensive ability in the "theory + practice" study.

Keywords

Learning management systems; Information extraction; Social networking sites.

1. Introduction

1.1. Motivation

In the era of big data, people are facing the challenge of a dramatic increase in the amount of data. Helpfulness analysis is the process of evaluating and judging the quality, credibility, relevance and other aspects of data, which can help people filter out valuable data and improve the efficiency of decision making. In the e-commerce domain, user reviews are an important source of information, but due to the huge number of reviews, how to extract useful information from them becomes a key problem. Therefore, we have built a Chinese comment dataset for the e-commerce field, hoping to help people perform helpfulness analysis through this dataset, and thus improve the service quality and user satisfaction. In addition, we can also use techniques such as sentiment analysis to understand users' sentiment tendencies towards products in order to better understand market demand and product optimization directions. By building this dataset, we can not only help e-commerce companies understand consumers' needs and improve product competitiveness, but also provide a data base for academic to promote research and development in related fields. In the future, we will continue to improve this

dataset and explore more application to provide more solutions for information processing in the era of big data.

According to the new engineering strategic concept proposed by the state, cultivating students' engineering practice ability and innovation ability become more important[1]. Ultimately reform of classroom teaching become the foothold in engineering education. The students need to master specific comprehensive experimental cases in addition to learning course expertise[2] when learning course of "Natural Language Processing". So, the traditional classroom teaching can no longer meet the demand for teaching. A flexible, Interactive, and convenient teaching platform can take the core of teaching to student-centered form teacher-centered and improve teaching across the board.

Some researchers have made useful explorations of natural language processing teaching or artificial intelligence, such as the reform of artificial intelligence technology[3], the approach of "scaffolded discovery learning"[4], the reform of OBE concept[5], distance learning[6,7], NLP4All[8] Platform, interdisciplinary thinking[9] and experimental teaching. Based on the actual situation and teaching requirements of computer science majors at China University of Mining and Technology (Beijing), this paper explores and develops the construction and application of the experimental platform based on Jupyter Notebook with the course "Natural Language Processing".

1.2. Introduction to Jupyter Notebook

The platform, based on Jupyter Notebook, allows students to interactively program in the form of browser remote access, supporting more than 40 programming languages including Python. Jupyter Notebook, which is often mentioned as a tool in the field of data science, is an open-source web application that facilitates the creation and sharing of program documents, supporting live code analyses, mathematical equations, visualization, etc.[10]. It can provide great convenience for the application and practice in the field of education. It consists of a series of cells, which comes mainly in two forms: code units and Markdown units. The code cells can Run the code directly, and Markdown cells can format text and insert links, pictures, and even mathematical formulas. Jupyter Notebook can meet almost all teaching needs. Students can also modify the Notebook according to their own learning habit. Classroom notes made through Notebook can be stored in the cloud and shared at any time. What's more, students can write and run programs directly on this platform.

2. Data Set

2.1. The need to build data sets

In e-commerce domain, user reviews evaluation is directly related to whether the buyer chooses to buy the product. There exists a strong relationship between a product's review and customers' willingness to buy it. Helpful reviews may provide the customers with a fast and convenient service while ineffective reviews will only take up customer's time and thus reduce purchase efficiency. Therefore, review helpfulness analysis will make both customers and platform the biggest beneficiary.

The existing criteria for evaluating effectiveness are not unified. Moreover, there is a lack of reviews comment validity score labels in Chinese. Therefore, this paper provides a validity scoring data set of Chinese review comments. For each comment, the following four aspects are scored: product description, service information, logistics information, and authenticity. The scoring criteria are divided into five levels, namely 2, 4, 6, 8 and 10. The more information about the description of the product contained in the review comment, the more the score, which is the amount of information contained. It does not mean that the better the evaluation, the higher the score in this aspect.

2.2. Rules

Validity score for four aspects are obtained by mode, the detailed rules for scoring the four aspects are as follows:

"Product description": the more information about the characteristics of the product itself, the higher the score. For example, the evaluation of a mobile phone, including battery life, charging time, standby time, camera pixels, mobile phone sound quality and other descriptions of commodity characteristics, the more the item information score.

"Service information": the characteristic description of communication with merchants, mainly the description of the seller's service attitude and the amount of information

"Logistics information": the seller's delivery speed and the description of the transportation time of the goods on the way. For example: "the delivery speed of

this store is very slow, and it took a week to deliver". Although this comment is a negative comment, the corresponding score in terms of logistics speed is very high.

"Authenticity" refers to the overall subjective judgment of the comment sentence. Compared with the above three, authentic comments are more subjective.

The volunteers scored in the data set were young people aged 20-30. The scores obtained according to the rules defined in this topic are shown in Figure 1.

Comment:	<u>第二次</u> 购买了。很 <mark>好用</mark> 。不伤耳朵。京东快递非常不错。京东快递值得			
(Earplugs)	信赖。			
	The second time. Very useful. It will not hurt your ears. Jingdong express is			
	very good. Jingdong express is trustworthy.			
Score:	Product description:4 Service information:2			
	Logistics Information:6 Authenticity:8			
(a)				
Comment:	鱼形很美很清晰,我买了 10 条,店主还多送两条备用的,只是有些鱼可			
(Goldfish)	能在运输途中把鱼鳞和鱼鳃给挂掉了!这个套餐鱼儿和饲料都有很划算!			
	值得推荐和购买。			
	The shape of the fish is very beautiful and clear. I bought 10, and the			
	shopkeeper gave two more for standby, but some fish probably hang off			
	their scales and gills during transportation! This set is very cost-effective for			
	fish and feed! Worth recommending and buying.			
Score:	Product description:10 Service information:6			
	Logistics Information:2 <u>Authenticity</u> :10			
	(0)			
Comment:	安装师傅非常专业、细心、开工之前做好防护措施、安全第一、提前告			
(Gree	知具体费用,安装的很好,很满意。宝贝收到了,颜色很好看, <u>质感</u> 也			
electric)	很好, <u>样子简单</u> !是我想要的!! <mark>下次</mark> 还会再来! <u>商品符合描述</u> ,主要是			
	活动便宜,能省钱,物流速度也很快,包装也很好			
	The installer is very professional and careful. He takes protective measures			
	before starting work, safety first, and informs the specific cost in advance. The			
	installation is very good and very satisfied.? Baby received, the color is very			
	good, the texture is also very good, the appearance is neat!? It's what I want!!			
	I'll come again next time! The product meets the description, mainly because			
	the activities are cheap which saves money, the logistics speed is also very			
	fast, and the packaging is also very good			
Score:	Product description:8 Service information:10			
	Logistics Information:6 <u>Authenticity</u> :2			

(c)

Figure 1: Comments and their scores

2.3. Detailed description of data set

Figure 2 shows the process of building a Jingdong commodity review dataset, which is roughly divided into the following steps:



Figure 2: The process of building a Jingdong commodity review dataset

Crawling Techniques: Using Web crawlers and crawler technologies such as Requests and Urlib3, a large amount of commodity user evaluation data is crawled from the Jingdong website to get raw data.

Data Processing: Preprocess the raw data, including data denoising, data de-weighting and other operations, to get clean data.

Evaluation validity criteria development: According to the research purpose and actual needs, a set of user review effectiveness evaluation criteria was designed, which includes 4 evaluation indexes: Product Description, Service Information, Logistics Information, Authenticity.

Data Analysis: According to the The data are manually labeled according to the designed evaluation criteria, i.e., each comment is scored according to the evaluation index, and a labeled data set is built.

Jingdong commodity review data statistics:9 categories, 114 items in total,102500 comments, 10000 marked. Table 1 shows the distribution of data categories.

3. Design of information processing experimental platform based on Jupyter Notebook in the context of "New Engineering"

3.1. Platform Architecture

In order to actively respond to the new round of industrial revolution, the Ministry of Education has been active in construction of new engineering to drive the construction of higher education powerhouse. The form of education in colleges must be able to foster new engineering talents and this paper constructs a Natural Language Processing experimental platform based on Jupyter Notebook. This platform breaks through the traditional teaching concepts[11], by overcoming timeliness and spatiality of communication in teaching. The platform is deployed on a server, and users can enter IP address in the browser to access the experimental platform.



Figure 3: Experimental platform framework diagram

Course modules	Knowledge	Cases
Household appliances 18 items 15000 comments, 2400 signed	Hisense TV (500) Compas alarm clock (1000) Deerma humidifier (1000) Table lamp (1000) Hisense TV (500,500 signed) Xiaomi household table lamp (500,500 signed) TV (500,150 signed) Water purifier (1000) Washing machine (1000)	FIYCO razor (1000) Konka refrigerator (1000) Midea hanging ironing machine(1000) TPLinkrouter (500,500 signed) Projector (500,500 signed) JBL wireless small sound (1000) Home audio system (1000) Sweeping robot (1000) Gree air conditioner(1000,250 signed)
Digital 16 items 15000 comments, 950 signed	Wacom digital board (1000) Edifier Bluetooth headset (1000) Xiaomi Bracelet (1000,200 signed) HP printer (1000) Victoriatourist computer bag (1000) Mobile phone film (1000) Laptop 2 (500) Blood glucose meter (1000)	Huawei stylus (1000) Glory power bank (1000,200 signed) Huawei smart screen (1000) Lenovo notebook (1000,200 signed) Huawei p50pro (1000,200 signed) Laptop 1 (1000) DJI UAV (1000,150 signed) ZMI power bank (500)
Household/furniture Kitchenware 11 items 10500 comments, 350 signed	Keep yoga mat (1000,250 signed) Thermos vacuum cup (1000) Byford mat (1000), Napkin (1000) Ergonomic chair (500,100 signed) Lifting table 2 (1000)	Spacexpert storage box (1000) RAE thermos cup (1000) Saiyi warm handbag (1000) Curtain (1000) Lifting table 1 (1000)
Clothing/luggage 20 items 17000 comments, 1650 signed	Anta casual sports shoes (1000) Tissot mechanical watch (1000) Earplugs (1000,200 signed) Scarf (1000) Paradise umbrella (500,500 signed) adivon (500) Freshman 2 (500) Skirt (500) Swan Necklace (1000) Down jacket 1 (1000)	Vantiiear suitcase (1000) Huixun umbrella (500) Ski goggles (1000) Luggage compartment (1000) Raincoat (500,500 signed) Coat 1 (1000,150 signed) Underwear (1000) Backpack (1000,150 signed) Sweater (1000,150 signed) Down jacket 2 (1000)
Beauty/cleaning 11 items 10000 comments, 1200 signed Sports, 3 items,	Pechoin frost (1000) LOREAL (1000,200 signed) Mask (1000) Perfume (1000) Lipstick (500,500 signed) Alcohol cotton sheet (1000) Phoenix mountain bike (1000)	Pechoin Facial Cleanser (1000) CIEAR Shampoo (1000) Hand sanitizer (1000) SK2 fairy water (500,500 signed) Foundation solution (1000) Basketball (1000)

Table 1: Selected course modules and corresponding cases

3000 comments	Table tennis racket (1000)	
	Wu Yutai jasmine tea (1000,200 signed)	
	Coconut milk (1000,200 signed)	Strawberry (1000,200 signed)
Food	Coke (1000)	Ferrero (1000)
15 items	Milk (1000)	Green tea (1000)
13500 comments	NESTLE Crispy shark (500, 150	Coconut (1000)
1400 signed	signed)	Spicy gluten (1000)
1400 Signed	Milk powder (1000)	Lemon (1000)
	Mangosteen (1000)	Raw chocolate (500,150 signed)
	Duck blood vermicelli soup (500,150 signed)	
	C language textbook (1000,200	Pen holder (1000)
	signed)	Globe (1000)
	M&G pen (1000)	Brocade flag (1000)
Stationery/toys,	Tuner (1000)	LEGO (1000)
Books,	English teaching volume (1000)	Romance of the Three Kingdoms
18 items,	Magic cube (1000)	(1000)
16500 comments,	Plasticine (1000)	Solitaire (1000)
1800 signed	Fang Siqi's first love paradise	Guitar (500,500 signed)
	(500,500 signed)	Pen (1000,100 signed)
	Trisomy (500,500 signed)	Breath of the wilderness
	Calculator (1000)	cassette(1000)
Pets/plants, 2 items,		E
2000 comments,250 signed	Goldfish (1000,250 signed)	Epipremnum aureum (1000)

The course on platform includes 10 chapters such as NLP overview, feature extraction, text machine learning, and 3 comprehensive project cases. In addition, each chapter has a chapter case to sum up the theoretical knowledge of this chapter. These contents are adhering to the principle from easy to difficult and the principle of gradual improvement. The platform architecture diagram is shown in Figure 3.

3.2. Design of platform function module

3.2.1. Online Learning Modules

The information processing experimental platform based on Jupyter Notebook break the bottleneck of traditional experimental courses in which Students practice only in the laboratory at the specified time. This platform is not limited by time and space, so that students can logging on to the platform at any time according to their actual situation to learn experimental courses, which can improve their learning efficiency.

After logging on to the platform, students can choose the course catalog independently, and deeply learn unfamiliar content, so as to consolidate their theoretical foundation. Students can use their own learning methods to mark the courseware on the platform and build knowledge system. They can also enter the server IP address and port to log in to the student-side background management system, so as to view the relevant information of the course. If the course has an examination, they can also inquire about the examination information and scores of the course. Online learning modules is beneficial to students' independent learning and improve the quality and efficiency of experimental teaching.

3.2.2. Interaction and Sharing Module

The platform has also developed a teacher-student interaction module, which allows students ask questions about doubtful knowledge or problems and difficulties encountered in the experimental process any time. Students also can discuss, communicate, and answer questions asked by other students. Teachers can also reply in time after seeing the questions. This can not only improve the quality of students' learning, but also effectively enhance the interaction between teachers and students.

After logging into the backend, students can click "My Classroom" - > "Exchange Q&A" in the main panel to enter the question interface, where the course, chapter, and title are required to fill out.

The sharing function of the experimental platform is also a major advantage. Teachers can import various resources when designing courseware. For simple experiments, teachers can import some knowledge links to allow students to learn independently and for complex programs, teachers can import external programs or provide corresponding program code to assist students to complete experiments. Besides, the platform also supports various formats to export documents in, such as PDF, HTML, Markdown, etc.

Teachers can share courseware and questions at any time, convenient for sharing, communication and discussion.

Through interactive sharing module, students can make progress in discussion and communication, and improve the quality of learning. And sharing module can also greatly enhance the interaction creating a good learning atmosphere.

3.2.3. Teaching Management Module

In addition to completing the basic teaching operation, teachers can log in the background of the system to manage the course, including course resource management, exercise management, examination management, statistical analysis, as shown in Figure 4. This Module build a complete online teaching resource library, which can assist experimental teaching conveniently and efficiently.



Figure 4: Teaching Management Module

(1) Curriculum resource Management.

Teachers can add, delete, or modify the lesson plans and chapters of the courses. The platform sets up a question bank. Teachers can set up question bank according to the course category, which can provide support for exercise data, and cooperate with the release of exercises, exams, and other scenarios.

(2) Exercise Management.

Teachers can release exercises of the lessons, and students can self-test through the exercises to identifying gaps and fill in. Teachers can learn about how well students master the knowledge and provide timely assistance to students who do not have a solid grasp of knowledge.

(3) Examination Management.

Teachers can use test to understand how students have mastered what they have studied in a certain stage, so that teachers can adjust the teaching in a planned or a targeted manner. The results of the stage examination can be summarized into the final score evaluation. Knowledge, ability and quality can be attached equal importance to the test content, so as to form a diversified assessment of students.

(4) Statistical analysis.

The platform offered a statistical analysis function to analyze students' class situation, exercises, exams and generate data. Teachers can view statistical analysis at any time. The platform also designed and developed online examination papers, which liberated teachers from traditional offline evaluation, reduced the burden, and increase productivity.

Compared with the traditional teaching management, the teaching management module makes teaching management simple and intelligent, which allows teachers to grasp the situation of students more deeply, adjust the teaching of follow-up courses in a timely manner, and improve the quality of teaching.

3.3. Natural Language Processing Professional Core Practice Course System Design

The design of the core practical course system of natural language processing courses takes practice-driven as the main line, including the theoretical knowledge of the course and the experimental design. The goal of the course is that students are familiar with common tasks in the field of natural language processing, master the general algorithm framework of natural language processing, and can apply natural language processing in different fields, so as to lay a good foundation for more indepth learning and work in related technical fields in the future.

In experimental teaching design of this course, it involves multiple knowledge such as semantic analysis, keyword extraction, word embeddings construction, text clustering, sentiment analysis and various machine learning algorithms, and the specific course experimental project design is shown in Table 2.

Text collection	Use Beautiful Soup 4, requests, Url1ib3, regular expressions, APIs, and other methods to collect text information, and extract information from JSON, XML format data	Online HTML page information extraction
Text preprocessing	Text segmentations, Part-of-speech annotation, Spelling correction, Lemmatization, Named entity recognition, Word Sense Disambiguation	Raw Text data preprocessing
Feature extraction	Keyword extraction, text data visualization (word cloud, named entity visualization, and dependency relation tree	Extract popular keywords from the news Text data visualization
Word Embedding	One-Hot coding,Keras Character encoding, Word2Vec,Doc2Vec	Document embedding news lookup
Sentiment analysis	Textblob basic sentiment analysis, SnowNLP Chinese text sentiment analysis,Emotional fine-grained analysis	textblob Sentiment analysis of social network texts
BERT model	BERT Text feature extraction,Masked LM(Close Test)	E-commerce user review classification prediction
Comprehensive cases	Data collection, text preprocessing and feature extraction, machine learning algorithm/deep learning algorithm implementation, model application	Capsule network fashion clothing pictures intelligent classification The community Q&A entity is precisely matched with emotion Attention Chinese-English Machine Translation

Table 2: Course experimental project design

4. Case Teaching on Experimental platform

The teaching feature of the platform is the combination of theoretical knowledge and practical application, and teachers can set relevant cases according to the content of each chapter, so that students can "learning by doing", enhance their practical ability, and deepen their understanding of theoretical knowledge. Case teaching, based on experimental platform, about the course content selects some basic knowledge of NLP theory for case teaching, and mainly expounds the three parts of Text collection, feature extraction and BERT model.

4.1. Text collection

Text collection is the first stage of NLP content, which need collect data according to the needs of the topic. The data source can be existing data, or collected data from various online repositories, or own data sets created through the crawlers.

In the Section of Collecting Online Text Data, the lessons teaches how to do text collection by using Requests and Urllib3 to get online text content of Project Gutenberg, a free e-book site, as an example. By importing the Requests library, sending the request to the specified URL, using the Requests object and citing the text attribute to display the text content of the obtained file, which can be saved to achieve online text data collection. Similarly, we can also use Urllib3 to implement the same operation, unlike the former, which need to import Urllib3 and generates a manager instance. The case text collection results are shown in Figure 5.

In [4]: r.text[:1000] # 显示前1000个字符

Out[4]: 'i» The Project Gutenberg EBook of David Copperfield, by Charles Dickens\r\n\r\nThis eBook is for the use of anyone anywhere at no cost and with\r\nalmost no restrictions whatsoever. You may copy it, give it away or\r\nre-use it under the terms of the Project Gutenberg License i ncluded\r\nwith this eBook or online at www.gutenberg.org\r\n\r\n\r\nTitle: David Copperfield\r\n\r\nAuthor: Charles Dickens\r\n\r\nRelease Date: December, 1996 [Etext #766]\r\nPosting Date: November 24, 2009\r\nLast Updated: September 25, 2016\r\n\r\nLanguage: English\r\n\r\nCh aracter set encoding: UTF-8\r\n\r\n*** START OF THIS PROJECT GUTENBERG EBOOK DAVID COPPERFIELD ***\r\n\r\n\r\n\r\n\r\n\r\n\r\n r\r\n\r\n\r\n\r\n\r\n\r\n\r\n\r\nDAVID COPPERFIELD\r\n\r\nBy Charles Dickens\r\n\r\n\r\n\r\n\r\n AFFECTIONATELY INSCRIBED TO\r\n THE HON. Mr. AND Mrs. RICHARD WATSON, \r\n OF ROCKINGHAM, NORTHAMPTONSHIRE. \r\n\r\nCONTENTS\r\n\r\n I Am B I Have a Change\r\n orn\r\n II. I Observe\r\n TTT. IV. I Fall into Disgrace\r\n V.

Figure 5: Text collection results

4.2. Feature extraction

Feature extraction is a very critical step in NLP, and the quality of feature extraction directly affects the subsequent research of applications. Computers can only understand binary numbers: 0 and 1, so every instruction we enter into a computer is converted to binary numbers, and likewise machine learning models tend to understand only numeric data. Therefore, you need to first convert the text data to its equivalent numeric form. Feature extraction is generally divided into two steps: word segmentation and word embedding construction. You can omit the word segmentation step if process English and focus on the teaching case of word embedding representation here.

Word embeddings research converts words and expressions into vector matrix as calculation models, and commonly used word representation methods include One-Hot, Word2Vec, etc. In the Manual Character Encoding section, we manually perform character-level One-Hot encoding by developing custom functions with simple examples, allowing students to experience the One-Hot encoding process and learn by doing it.

First define the encoding function, use the enumerate() function to convert the object to dictionaries. After that, create full 0 vectors of the same length as the dictionary, traversal each char, make the characters match the dictionaries one by one, set the match-consistent index to 1, and keep the rest at 0, completing the One-Hot encoding; Then perform the One-Hot encoding test, adding the word "data" as the input to the function and check the results. The specific code is shown in Figure 6.

Figure 6: One-hot Encoding case

According to the index of the tag encoding, the One-Hot encoding of letter 'd' is [1,0,0], the letter 't' is encoded as [0,1,0], and the letter 'a' is encoded as [0,0,1], and for illustrative convenience, the dictionary length is set to 3 in this case.

4.3. BERT model

Bert Pre-Training Model is a large multi-task language model based on the attention algorithm proposed by Google, which has been widely applied to NLP tasks, and has achieved breakthrough results in eleven directions in the field of NLP, which is another epoch-making model, which achieves accuracy improvement after Word2Vec. It is characterized by the use of transformer structures that have advantage of capture of bidirectional relationships in text, Language Model and Predict the Next Sentence multi-task learning.

In the chapter of the BERT model, this teaching designed the e-commerce user review classification prediction experiment. According to the user's review text of the product, realized the cost performance, product quality, participation in activities, customer service and logistics packaging, whether it is an old customer, whether it will buy again, the overall review seven dimensions of the classification prediction, and output positive, negative or neutral three emotional predictions for each dimension, the test results are shown in Figure 7.



Figure 7: Outcome of the BERT model predicts

4.4. Comprehensive cases

In addition to the cases in each chapter, this curriculum also designs comprehensive cases, allowing students to quickly get started and deeply experience the development process of NLP, and cultivate students' manual practical ability. This article takes the case of "intelligent classification of capsule network fashion clothing pictures" as an example to conduct an indepth discussion.

This case implements the capsule network architecture though Keras and TensorFlow backends. Based on the model framework of the capsule network, node calculation and loss function setting which has taught in the classroom, this case helps students to understand the core ideas of capsule networks and strengthen students' understanding of the limitations of convolutional neural networks. At the same time, a new generation of deep learning benchmark dataset Fashion MNIST is introduced, and the model training and evaluation of capsule network architecture are implemented on this dataset. The specific case flow chart is shown in Figure 8.



Figure 8: Capsule network fashine clothing picture intelligent classification case flow chart

Through the study of this case, students will establish a preliminary understanding and realization foundation of the capsule network.

The teaching method of combining theory and practice allows students to learn to apply the theoretical knowledge they have learned to practical application projects, improve their manual practical ability, and solve the problem of students "learning without knowing how to use".

5. Conclusion

This paper uses the "Natural Language Processing" course as the carrier to build an experimental platform based on Jupyter Notebook. The teaching content and the arrangement of experimental courses is reformed and explored, and based on this, the experimental project cases of the course are designed, and from basic knowledge theory to complex project practice, the ability of students' natural language processing comprehensively gets comprehensive exercise.We also use the constructed dataset as experimental data on the platform. The dataset covers multiple topics, and has high quality and representativeness, which can help students conduct experiments on various tasks such as text classification, helpfulness analysis, keyword extraction and so on.

This platform provides a lightweight and convenient natural language processing teaching environment, the platform has been a trial run of experimental teaching for students majoring in computer science of China University of Mining and Technology (Beijing). Teaching practice shows that the application of this platform can greatly enhance students' enthusiasm for the curriculum and lay a good foundation for students to participate in important competitions such as the Blue Bridge Cup and ACM programming competition. In addition, students can also use this platform to complete the experiments of the course at home, which ensures the efficiency of students' online learning and strongly assist the experimental teaching[12]. We believe that our constructed dataset can not only be used for teaching purposes, but also for research and application in other fields, contributing to the NLP field.

References

- [1] Xiren Li: Opportunity and Challenge of the Cultivation of Talents Under the Background of Artificial Intelligence. International Conference on Advances in Energy, Environment and Chemical Science (Zhuhai, China, February 2-4, 2018). Vol. 189, p.350-353.
- [2] Yujun Zhou, Zehui Zhan, Lu Liu, et al: International Prospects and Trends of Artificial Intelligence Education: A Content Analysis of Top-level AI Curriculum across Countries. Proceedings of the 6th International Conference on Digital Technology in Education (Beijing, China, September 16-18,2022). Vol.6, p.337-343.
- [3] Mehrnaz Fahimirad, Sedigheh Shakib Kotamjani, et al: A review on application of artificial intelligence in teaching and learning in educational contexts. International Journal of Learning and Development, Vol. 8 (2018) No.4, p.106-118.
- [4] Alexandra Schofield, Richard Wicentowski, Julie Medero: Learning How To Learn NLP: Developing Introductory Concepts Through Scaffolded Discovery. Proceedings of the Fifth Workshop on Teaching NLP. Association for Computational Linguistics (Online, June 11, 2021), Vol. 1, p.131–137.
- [5] Yuexia Tang, Guangping Li: Research and Practice of PLC Experiment Teaching Based on OBE and Digital Twin. Advances in Artificial Systems for Medicine and Education VI (Beijing, China, January 21-23, 2022), Vol. 159, p.359–368.
- [6] Miaomiao Jia: The Influence of Distance Learning during COVID-19 Pandemic on Student's Self-Regulated Learning in Higher Education: A Qualitative Study. Proceedings of the 5th International Conference on Digital Technology in Education (Busan, Republic of Korea, September 15-17, 2021), Vol. 6, p.47-52.
- [7] Zhou Jin, Yingxin Li, Chien-Liang Lin: The Effectiveness of Information Technology and Teaching Strategies in Distance Learning to Improve the Learning Effect of K-12 Students: Instructional Strategies and Application of Information Technology in Distance Learning. Proceedings of the 6th International Conference on Digital Technology in Education (Beijing, China, September 16-18,2022). Vol. 7, p.186–192.
- [8] Rebekah Baglini, Arthur Hjorth: Natural Language Processing 4 All (NLP4All): A New Online Platform for Teaching and Learning NLP Concepts. arXiv preprint arXiv:2105.13704.
- [9] Anoop Mishra, Harvey Siy: An interdisciplinary approach for teaching artificial intelligence to computer science students. Proceedings of the 21st annual conference on information technology education. (Online, October 7-10, 2020), Vol. 1, p.344–344.
- [10] Hélia Guerra, Luís Mendes Gomes, Alberto Cardoso: Agile approach to a CS2-based course using the Jupyter notebook in lab classes. 2019 5th Experiment International Conference (Funchal, Portugal, June 12-14). Vol. 19, p.177–182.
- [11] Alberto Cardoso, Joaquim Leitao, César Teixeira: Using the Jupyter notebook as a tool to support the teaching and learning processes in engineering courses. The Challenges of the Digital Transformation in Education: Proceedings of the 21st International Conference on Interactive Collaborative Learning (Kos Island, Greece, September 25-28). Vol. 2, p.227–236.
- [12] Christine Heitz, Martha Laboissiere, Saurabh Sanghvi, et al: Getting the next phase of remote learning right in higher education. McKinsey Insights (2020).