



Graduation Thesis Topic Recommendation Based on Neural Network

Xuemei Li^{1, a*}, Maryli F Rosas^{2, b}

¹College of Science and Computer Studies – Graduate Studies De La Salle University –Dasmariñas City of Dasmariñas, Cavite, Philippines

²College of Science and Computer Studies – Graduate Studies De La Salle University –Dasmariñas City of Dasmariñas, Cavite, Philippines

^alxx2677@dlsud.edu.ph

^bmfrosas@dlsud.edu.ph

Abstract

Undergraduate graduation design thesis is not only the goal of higher education, but also an important link. It plays an important role in training college students to carry out scientific research, basic training, and improve their ability and comprehensive quality. This paper takes the engineering class of the College of Electronic and Electrical Engineering as the research object, the paper uses text mining algorithm to know the main research topic and the simulation software used. In order to solve the problem of how to guide the students to choose the suitable graduation project theme, this paper puts forward some suggestions to the students through the method of neural network.

Keywords: text mining, neural network, thesis topic recommendation

1 INTRODUCTION

Although different majors in different universities may have different requirements for graduation design, this paper only takes the electrical engineering and automation majors of science and engineering as an example to mine the data of bachelor's degree thesis in the past three years. Because in the annual graduation design thesis, the most difficult thing is the topic selection of students or teachers. The thesis requirement for a bachelor's degree is that there can be no repetition of topics in the past three years. Because young teachers lack experience, the range of knowledge involved is limited, and students do not know enough about the scope of knowledge involved in this major, they often feel at a loss when choosing topics. At the same time, students may not know what factors their final paper scores are related to, or at least do not know exactly which factors are related to them. Students sometimes have many factors to consider when choosing a thesis topic, such as the difficulty of knowledge points, personal interests and good subject software.

This paper wants to use text mining to let teachers and students know more about the scope and direction of choosing topics for papers in this major. At the same time, we can know which topics are more popular through data

mining, and we can avoid repeated topics. The ultimate purpose of this paper is as follows: 1. It is relatively easy for students to understand which topics are related to knowledge points. This is mainly used as a reference for teachers to teach students in accordance with their aptitude. 2. Students learn what kind of thesis they would be good at.

2 RELATED WORK

A threshold electromagnetic classification approach for cylinders embedded in a lossy medium by using a neural network technique [3]. In this paper, a method of threshold classification is proposed. This method is used to distinguish conductors embedded in lossy media from conducting cylindrical objects. On the basis of the neural network, this method evaluates the conductivity of the embedded cylinder from the scattered electric field near the illuminating source, and the final result is that it provides a high percentage of successful classification in the presence of noise.

Comparative Research on GPS Height Fitting Methods based on Neural Network, [8]. They proposed a method of how to convert the GPS geodetic height into the actual accuracy of the normal height. By comparing and discussing the height of GPS under three different

conditions, these three conditions are application of BP, genetic and annealing neural networks. Finally, it is concluded that BP neural network has the ability of nonlinear mapping, and it is feasible to carry out high fitting. The accuracy is also higher than that of conical fitting. Genetic neural network integration is better than BP neural network, and the speed is faster. Simulated annealing neural network is an optimization of BP, which improves the accuracy.

Application of Neural Networks in Computer Security [6]. In this paper, he has several purposes. They first analyze the security of communication in the control system, then determine an appropriate type of neural network through data transmission verification, and propose a functional security system model. This model has been supported by experiments of various architecture types and neural network activation functions, and then tested in the real environment. Finally, a functional system scheme with practical application value is put forward.

Time series forecasting using artificial neural networks methodologies: A systematic review, [1] This is a literature review on the research of neural network method on time series prediction model. This paper uses various methods to predict the time series of the models related to neural networks in the papers published in the past decade, and summarizes them. Most literatures have proposed the application of neural network model, but there are few new neural network models, so it is very important to establish new neural network models.

A text mining application on monthly price developments reports [4], text mining analysis provides a great opportunity for economic research based on language processing technology, allowing us to read multilingual and national central bank price development reports. This article on economic text analysis analyzes the text, and then finds out that themes and clusters include some emotions and psychology. The study found that the assessment of inflation included some reinforced references to core groups and their interest in inflationary trends rather than consumption levels.

A survey of the applications of text mining for agriculture [2], This paper is based on text mining of agricultural literature, such as scientific papers and news reports, which can be analyzed by text mining technology to solve agricultural problems and extract some valuable knowledge. these knowledge are very potential, but text mining is rarely used in the field of agriculture. Therefore, this study provides an application of text-based mining technology in agricultural problems.

3 CHAIN STORE LOCATION MODEL

The research study employs the quantitative method of research. In this study, we will use the process of KDD (Knowledge Discovery in Database). This process is

a process of collecting data, processing the data through mathematical statistics, getting a conclusion, and finally transforming it into knowledge. There are two data mining methods used in this paper, including text mining, Neural network.

3.1. Steps of a KDD Process

Knowledge Discovery (KDD) is a nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in Databases. Knowledge discovery turns information into knowledge and finds the gold nugget of knowledge from the data mine, which will contribute to the development of knowledge innovation and knowledge economy [5].

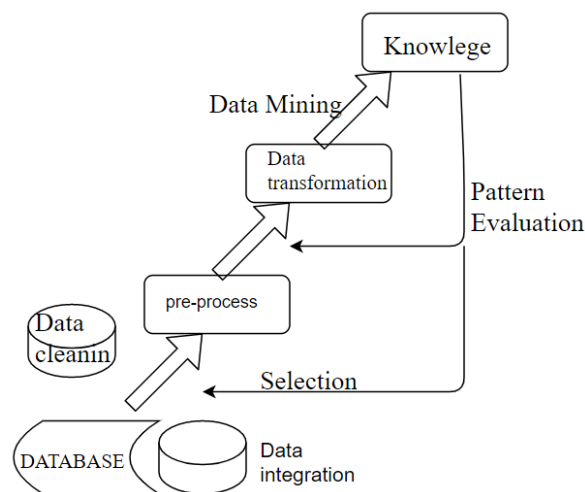


Figure 1 Steps of a KDD Process

1). Selection: In the first step we often need to know what data can be applied to our KDD project. The first step is the selection and collection of data. This paper chooses the master's degree thesis of undergraduate students majoring in electrical engineering and automation in the past two years. The method is to put all the abstracts of the master's degree thesis in one text, which is convenient for text mining.

2). Pre-processing: After the data is collected, the next step must be to pre-process the data and eliminate the existence in the data as far as possible error and missing information. In this step of data preprocessing, some irrelevant data attributes are deleted. For example, the name of the student, the corresponding instructor and other irrelevant attributes have been deleted.

3). Transformation: Transform the data into the format required by the data mining tool. This step can make the result more ideal. In this step,

4). Data mining: Apply data mining tools. Choosing the mining algorithm(s). There are two data mining methods used in this paper, including text mining, Neural network.

5). Interpretation/evaluation: The last step This step is to adjust our model through evaluation to improve the accuracy of the model. The main purpose of this step is to optimize the model.

3.2 Text mining

Text mining is to obtain meaningful data and information from a large number of words and data. Text mining is the process of obtaining patterns of interest to users from a large amount of text information [7]. The data statistics method of text mining can be classified and clustered. Classification is supervised learning, clustering is unsupervised learning.

Text knowledge discovery mainly consists of the following steps:

1). Text preprocessing: select task-related text and transform it into an intermediate form that can be processed by text mining tools.

2). Text mining: after the completion of text preprocessing, methods such as machine learning, data mining and pattern recognition can be used to extract knowledge or patterns oriented to specific application targets.

3). Model evaluation and representation is the last step, which uses the defined evaluation indicators to evaluate the acquired knowledge or model. If the evaluation results meet the requirements, the model is stored for use by users; Otherwise, go back to the previous part of the process to re-tune and improve, and then make another round of discovery.

3.3 Artificial neural network

Artificial neural network is a nonlinear dynamic system, which is characterized by distributed information storage and parallel cooperative processing. Although the structure of a single neuron is extremely simple and the function is limited, the behavior of a network system composed of a large number of neurons is extremely colorful.

The purpose of the neural network algorithm in this paper is to recommend students' thesis topics in different directions according to their good software, skills and achievements.

4 EXPERIMENT

In this paper, a total of two data mining methods have been done.

4.1 text mining

The first kind is text mining, which aims to provide a suggested guidance for students who are preparing to choose the subject of their thesis, and finally help students

to pass the defense of graduation thesis smoothly. The data we use are more than 200 student degree theses on electrical engineering and automation in the last two years. There are 7 folders in total. The result is shown in the following Table1.

Table1. Word frequency

word	Attribute name	Total Occurrences	Document Occurrences
Of-the	Of-the	473	6
control	control	470	5
system	system	388	6
a	a	342	6
design	design	277	6
power	power	222	5
circuit	circuit	200	5
can	can	199	6
on	on	197	6
paper	paper	118	6
speed	speed	117	5
Algorithm	Algorithm	112	4
Electric	Electric	103	5
simulate	simulate	93	6
control	control	90	5

From Table1, after removing some irrelevant words, circuit design or power, control systems, these vocabularies appear more frequently.

The frequency of algorithm or simulation is also higher. In other words, in the major of electrical engineering and automation, the thesis topic that most students can think of is generally power control system, which should occupy most of the topics in the thesis.

Next, take a look at the software mainly used in this major as keywords and take a look at the frequency of these keywords. As shown in Table2 as follows.

Table2 Frequency of Matlab

word	Attribute name	Total Occurrences	Document Occurrences
Com bin	Com bin	22	5
design	design	22	3
good	good	22	6
input	input	22	4
loop	loop	22	3
Matlab	Matlab	22	5

Table2 shows that the word Matlab appears 22 times, it also shows that there are seven topic selection groups, and five groups have chosen Matlab as the topic selection direction of the thesis. Therefore, Matlab software should be a mainstream software in the major of electrical engineering automation.

4.2 Artificial neural network

This data comes from the scores of students in three classes at a certain level of electrical and automation and the software they are good at. The result of the Artificial neural network is shown below:

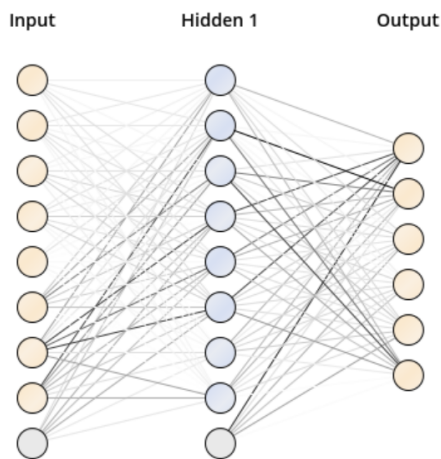


Fig2. Neural network results of thesis topic

Because there are too many nodes in the figure, we only show the results of some nodes, as follows:

Hidden 1

Node 1 (Sigmoid)

Sensor and detection technology: -0.257

Automatic control theory: 0.415

Power electronics technology: -0.269

Power supply and distribution technology: -0.346

Electrical control and PLC technology: 0.166

Automatic control system for electric drive: 0.923

auto CAD/Matlab: 1.656

Pro tel/FPGA/DSP: -0.038

Bias: -0.635

Node 2 (Sigmoid)

Sensor and detection technology: 0.443

Automatic control theory: 0.403

Power electronics technology: 0.150

Power supply and distribution technology: 0.521

Electrical control and PLC technology: 0.507

Automatic control system for electric drive: -1.536

auto CAD/Matlab: 0.600

Pro tel/FPGA/DSP: -3.179

Bias: 1.079

Node 3 (Sigmoid)

Sensor and detection technology: -0.168

Automatic control theory: 0.221

Power electronics technology: 0.492

Power supply and distribution technology: -0.333

Electrical control and PLC technology: 0.126

Automatic control system for electric drive: 2.190

auto CAD/Matlab: -0.399

Pro tel/FPGA/DSP: 1.807

Bias: -0.032

Node 4 (Sigmoid)

Sensor and detection technology: -0.555

Automatic control theory: 0.543

Power electronics technology: -0.697

Power supply and distribution technology: -0.770

Electrical control and PLC technology: 0.026

Automatic control system for electric drive: 0.919

auto CAD/Matlab: 3.361

Pro tel/FPGA/DSP: -1.150

Bias: -1.432

Node 5 (Sigmoid)

Sensor and detection technology: -0.400

Automatic control theory: 0.453

Power electronics technology: -0.524

Power supply and distribution technology: -0.537

Electrical control and PLC technology: 0.086

Automatic control system for electric drive: 0.806

auto CAD/Matlab: 2.649

Pro tel/FPGA/DSP: -1.353

Bias: -1.169

Node 6 (Sigmoid)

Sensor and detection technology: -0.571
 Automatic control theory: 0.511
 Power electronics technology: -0.753
 Power supply and distribution technology: -0.786
 Electrical control and PLC technology: 0.091
 Automatic control system for electric drive: 0.978
 auto CAD/Matlab: 3.398
 Pro tel/FPGA/DSP: -0.943
 Bias: -1.464

Node 7 (Sigmoid)

Sensor and detection technology: -0.033
 Automatic control theory: -0.254
 Power electronics technology: 0.049
 Power supply and distribution technology: 0.167
 Electrical control and PLC technology: -0.158
 Automatic control system for electric drive: 0.027
 auto CAD/Matlab: -0.494
 Pro tel/FPGA/DSP: 0.575
 Bias: 0.418

Node 8 (Sigmoid)

Sensor and detection technology: -0.202
 Automatic control theory: 0.171
 Power electronics technology: -0.314
 Power supply and distribution technology: -0.187
 Electrical control and PLC technology: -0.056
 Automatic control system for electric drive: -0.012
 auto CAD/Matlab: 1.620
 Pro tel/FPGA/DSP: -1.354
 Bias: -0.437

In addition, we can also get some tables, this table is only for the choice of thesis topic as a reference. Some of the results are shown below. Software simulation(S-S), Hardware design system (HDS), Plant power supply design (PPSD), New energy research (NER).

Table 3 Graduation thesis topic recommendation

Prediction (topic of thesis)	Confidence(S-S)	Confidence (HDS)	Confidence (PPSD)
HDS	0.011	0.717	0.043
S-S	0.901	0.003	0.012
NER	0.066	0.107	0.069
HDS	0.012	0.765	0.038
S-S	0.093	0.003	0.01
HDS	0.01	0.784	0.03

As we can see from Table 3, according to the students' different personalities and the different software they are good at, as well as the different scores of each subject, we can use the neural network algorithm to provide students with some recommended choices. For example, if a student is good at hardware circuit design, we will recommend him to do some graduation thesis topics related to hardware circuit design, perhaps using software Pro tel. However, we do not have such absolute results, if there is a suggested result for students to choose their own graduation thesis topic.

5 CONCLUSION

There are still many shortcomings. First, we can expand the scope of students. Second, more personal factors can be collected, such as independent variables, in order to determine the theme direction of students' graduation thesis more accurately.

The conclusion of this paper shows that the main topics of the dissertation of Electrical Engineering and Automation are control, system design and simulation. The main software used is Matlab. When we advocate the diversity of thesis topics, we should try to avoid using Matlab as the only topic. The results of neural network show that students can get different recommendations of graduation thesis themes according to their own different factors, which plays a constructive role in students' choice of graduation thesis direction and teachers' guidance to students.

ACKNOWLEDGMENT

I would like to thank Li afang, a teacher from the School of Electronic and Electrical Engineering, Anhui Sanlian University, for the data provided.

I would like to thank Maryli F.Rosas professor for her careful guidance.

PROJECT INFORMATION

1. Anhui Provincial teaching and research project name: Electric power drag automatic control system DC speed regulation system basic concept wisdom classroom pilot (No.: 2017zhkt)

2. Anhui Sanlian University Quality Engineering Project "Roadbed and Pavement Engineering Intelligent Classroom Pilot" (No.: 18zlgc057)

3. Anhui Provincial Quality Engineering Project "Electronic and Electrical Experimental Training Center" (No.: 2018sxzx54)

4. Research project of Anhui Sanlian University "Research on reactive power and Harmonic Problem of AC and DC Side under Full Control Switching Device Rectifier System" (No. PTZD2020006)

REFERENCES

- [1] Ahmed Tealab, 2018. Time series forecasting using artificial neural networks methodologies: A systematic review. ScienceDirect, Future Computing and Informatics Journal 3, 334e340 <http://www.journals.elsevier.com/future-computing-and-informatics-journal/>
- [2] Brett Drury, Mathieu Roche. 2019. A survey of the applications of text mining for agriculture, Contents lists available at ScienceDirect, Computers and Electronics in Agriculture 163 (2019) 104864
- [3] E. Bermiani, S. Caorsi, 1 and M. Raffetto, 2000. A threshold eletromagnetic classification approach for cylinders is embedded in a lossy medium by using a neural network technique. MICROWAVE AND OPTICAL TECHNOLOGY LETTERS / Vol. 24, No. 1, January 5.
- [4] Hatice Burcu Eskici, Necmettin Alpay Koçak, 2018, A text mining application on monthly price developments reports, Contents lists available at ScienceDirect Central Bank Review 18, 51-60,
- [5] Karamshuk D, Noulas A, Scellato S, et al. 2013. Geo-spotting: mining online location-based services for optimal retail store placement [C] // Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and datamining. Chicago: ACM: 793–801.
- [6] Halenar Igor, Juhasova Bohuslava, Juhas Martin, Nesticky Martin, 2014. Application of Neural Networks is used in Computer Security, ScienceDirect, Procedia Engineering 69, 1209 – 1215, Available online at www.sciencedirect.com
- [7] Liy, Zheng Y, Ji S, et al. 2015. Location selection for ambulance stations: a data -driven approach [C]// Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems. [S.l.]: ACM, 85-91.
- [8] NIU Zhihong, SONG Mengbo, CHEN Jiqin. 2012. Comparative Research on GPS Height Fitting Methods based on Neural Network. Sciverse science Direct Procedia Engineering 28, 696 – 702. Available online at www.sciencedirect.com

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

