Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD27290

A Major Recommendation System for University Admissions Based on Knowledge Graphs and Large Language Models

Yutong Leng

Institute of Software Engineering, Jilin University, Changchun, China
13756987653@163.com

Abstract. The majority of college entrance examination candidates in China come across substantial difficulties and experience confusion during the process of filling out their preferences for universities and majors. The existing systems that offer access to historical admission data can give some support in the decision - making process. Nevertheless, these systems usually demand that users input information in a rigid and structured format. This makes it quite hard to provide support for natural and personalized interactions with students. On the contrary, large language models (LLMs) have powerful capabilities in understanding and generating natural language, which can bring about more user - friendly dialogue interfaces. However, because of the well - known problem of hallucination in LLMs, the reliability and factual correctness of the outputs they generate cannot be fully ensured. To deal with these challenges, this paper constructs a structured knowledge graph of university - major admission data on the basis of historical college entrance examination scores. This knowledge graph acts as a trustworthy information base to carry out the process of constraining and enhancing the recommendation process. Based on this, we put forward a university and major recommendation framework that integrates extraction - enhanced techniques and large language models. By making use of the knowledge graph to guide the generation process, our system can improve both the accuracy and the relevance of the recommendations. The experimental results show that the proposed method can significantly enhance the precision of university and major recommendations, which indicates that it has strong potential for real - world application and further development.

Keywords: Major Recommendation, Knowledge Graph, Large Language Model, College Admissions, and Recommendation System

1. Introduction

In China, college entrance examination candidates, when they are in the process of selecting universities and majors, often go through a state of confusion and anxiety. According to Cheng et al. [1], more than 71% of senior high school students have a lack of the ability to carry out an assessment of the alignment between their interests and potential majors, and around 66% are not certain about their future career paths or how these paths relate to academic disciplines. Even though artificial intelligence (AI)-based systems have been developed to offer support for decision - making, in most families, these tools are mainly operated by parents, and there is limited involvement of students. Existing systems have a tendency to put emphasis on score - based matching, while they overlook long - term career alignment and personal interest orientation [2]. Traditional recommenders have a

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD27290

preference for popular items, and this may strengthen echo chambers and ignore users' less dominant interests or emergent preferences [3]. Moreover, studies show that various external influences, such as family expectations, school guidance, university promotion strategies, and online platforms, play a significant role in the process of shaping students' decisions when there are conflicting factors. These conflicting factors include interest versus return on investment, idealism versus employment reality, and societal expectation versus personal uncertainty [4].

In the situation where artificial intelligence is advancing rapidly, large language models (LLMs) have come out as powerful tools for the process of natural language interaction and the process of information processing. However, a critical limitation of these large language models lies in the fact that they are susceptible to the situation of hallucination, that is to say, they generate content that is factually incorrect or fabricated. Adel et al. [5] report that in the process when generative AI tools are used, they can reduce the workload of tasks by 60 to 65% on average. But their performance in tasks that require the process of explanation or the process of precise reasoning remains extremely low. The accuracy drops to as little as 4.6% and the rates of hallucination reach up to 91%. Theoretical studies further put forward the argument that large language models are fundamentally not capable of learning all computable functions, which makes hallucination a phenomenon that is mathematically inevitable [6]. Even among leading models, for example ChatGPT, this kind of problem still exists. GPT - 3.5 shows a hallucination rate of 39.6%, GPT - 4 shows a rate of 28.6%, and Google Bard shows a staggering rate of 91.4% [7].

To tackle the above - mentioned problems of interest mismatch and hallucination, in this study, a novel recommendation system is put forward. This system is based on a structured knowledge graph. This knowledge graph is constructed from historical university admission data. The process of constructing this knowledge graph involves taking into account historical university admission data.

The knowledge graph integrates key entities like universities, majors, regions, scores, and enrollment years. That is to say, it combines these key entities such as universities, majors, regions, scores, and enrollment years together, and this provides a reliable basis for making well - informed recommendations.

On top of this graph, we carry out the development of a recommendation framework. This framework is enhanced with an information extraction mechanism. This information extraction mechanism guides the large language model so that it can generate more accurate and contextually relevant suggestions.

The experimental results show that our approach significantly outperforms traditional score - based systems when it comes to the aspect of recommendation accuracy. This indicates its practical value and the potential for broader deployment.

The overall system architecture is presented in Figure 1. In other words, the whole system architecture is shown in Figure 1, which gives a clear picture of the system's structure.

DOI: 10.54254/2755-2721/2025.LD27290

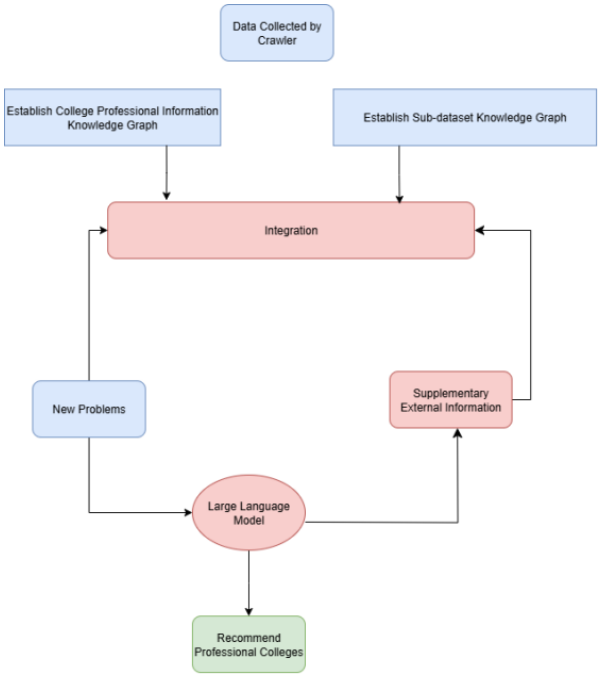


Figure 1. Overall system architecture

The remaining part of this paper is arranged in the following way: In Section 2, an introduction is given to the core technologies that are employed in our system. In other words, it is about the process of introducing the core technologies used in our system. Section 3 goes into details about the implementation framework. This includes the process of constructing the knowledge graph and the process of setting up the enhanced extraction mechanism. Section 4 presents the experimental setup and also shows the evaluation results. That is to say, it is about presenting the experimental setup and then presenting the evaluation results. Section 5 summarizes the key findings of this study. In other words, it is about carrying out the process of summarizing the key findings of the study. And Section 6 discusses the potential directions for future research and the ways to improve the system. That is, it is about having a discussion on the potential directions for future research and then discussing the ways to make improvements to the system.

2. Related work

2.1. Applications of Knowledge Graphs

The concept of the Knowledge Graph (KG) was officially brought forward by Google on the seventeenth day of May in the year 2012. This was done as an attempt to carry out an enhancement of the capabilities of search engines. The way to achieve this enhancement was by means of incorporating structured semantic information into the search engines. Nevertheless, the technological foundation of it can be traced back to the situation where Berners - Lee put forward the Linked Data paradigm in 2006. This paradigm had the vision of making an interconnection of distributed web data through the use of standardized protocols. A knowledge graph usually acts as an auxiliary knowledge base. This knowledge base undertakes the task of integrating entities and their relationships. By doing so, it enables the implementation of semantic - level computation and reasoning.

The core technologies that serve as the foundation for the process of knowledge graph construction encompass knowledge extraction, knowledge fusion, as well as knowledge refinement. In the process of knowledge extraction, it involves the act of identifying entities and the relationships between them

DOI: 10.54254/2755-2721/2025.LD27290

from data sources that are either unstructured or semi - structured. When it comes to knowledge fusion, it is about dealing with the situation of resolving conflicts and redundancies that exist across heterogeneous sources. And for knowledge refinement, it is to ensure the consistency and usability of the knowledge graph through carrying out schema alignment and implementing quality control.

In the process when the era of big data and the semantic web are advancing, knowledge graphs have come into being as the foundational infrastructure that is applicable to a wide array of applications. The contributions that they make to downstream tasks, such as information visualization, question answering, intelligent summarization as well as semantic search, have been widely validated in the situations where both academic and industrial activities are carried out [8]. Contemporary AI systems, including those prominent large - scale models like ChatGPT and DeepSeek, have, as time goes on, increasingly incorporated knowledge graphs so as to carry out the mitigation of hallucinations and let the factual correctness of generated outputs be enhanced.

Empirical studies have carried out a further highlighting of the pedagogical potential that knowledge graphs possess. For example, in the work done by Yang et al. [9], it is demonstrated that when one integrates knowledge graphs into educational recommendation tools, it not only brings about an improvement in the precision of content suggestions but also promotes the engagement of students and enriches the interactive learning experiences within the process of online environments. These kinds of findings emphasize the multifaceted utility of knowledge graphs in both the area of cognitive computing and the applications that are centered around human beings.

2.2. Applications of Large Language Models

Large Language Models (LLMs) stand for a category of deep - learning architectures that are designed in such a way that they can understand, generate, and manipulate human language on a large scale. They depend on a set of enabling technologies. These technologies include scaling laws, data engineering, efficient pretraining strategies, capability elicitation, human alignment, and tool integration. In the process of showing their capabilities, the unprecedented depth of linguistic and conceptual understanding that LLMs display makes them different from traditional natural language processing models. This also indicates a significant advancement in AI capabilities.

Large Language Models (LLMs) have shown a high - level of strong performance when it comes to a diverse set of domains. In the process of considering the educational sector, they are extensively and widely applied to carry out tasks like exam question generation, automated grading, personalized feedback, as well as adaptive content recommendation. All these tasks, taken together, encompass a total of nine distinct categories of educational applications [10].

When it comes to the software engineering field, LLMs offer support for code generation, bug fixing, code explanation, and natural language documentation. By doing so, they are able to accelerate the software development life cycle and also lower the technical barrier that exists for non - expert programmers [11].

In the situation of looking at the financial domain, LLMs have the ability to perform zero - shot and few - shot learning, fine - tuning, and the construction of customized models. A decision - making framework that is specifically tailored for financial professionals has been put forward, which highlights the associated challenges and limitations [12].

The more extensive scientific community has also taken on large language models (LLMs) as tools for making discoveries. In the situation where Zhang and others carried out a comprehensive analysis, they analyzed more than 260 large language models that are customized for scientific use - cases. They documented the process of integrating these large language models into cross - modal tasks in the fields of chemistry, biology, astronomy, as well as climate science [13]. The findings they obtained show the ability of large language models to promote the process of hypothesis generation,

DOI: 10.54254/2755-2721/2025.LD27290

to carry out the process of data interpretation, and to perform the process of automated reasoning. By doing so, they can make the scientific research pipeline more efficient.

Overall, in the situation where large language models (LLMs) are being deployed in a widespread manner across both technical and non - technical domains, this has made them become transformative tools in the field of artificial intelligence. They have demonstrated their capabilities in the process of various fields, including education, medicine, legal analysis, and scientific discovery [14]. Moreover, in the process of their integration with structured knowledge sources, such as knowledge graphs, it has been shown that this can significantly enhance the factual reliability, the interpretability, as well as the performance in real - world applications. This is especially the case in the process of mitigating hallucinations and supporting semantic - level reasoning [15].

3. Methodology

3.1. Knowledge Graph construction

This study undertakes the process of constructing two types of domain - specific knowledge graphs in order to provide support for university and major recommendation. These two types are: (1) the graph which can be described as the "Major - University - Admission Score Graph" and (2) the graph named the "Major - Interest Graph". In the case where the "Major - University - Admission Score Graph" exists, it makes it possible to carry out decision modeling for college entrance recommendation. On the other hand, the "Major - Interest Graph" offers support for the process of major personalization that is based on the individual's interests.

3.1.1. Major-university-admission score graph

This graph is made up of two sub - structures. First, there is (a) the layer of the university base information. Second, there is (b) the records of the 2024 admission scores for each major in each university in Liaoning Province, China. The base information of the university includes attributes like the name of the university, the province where it is located, the city where it is situated, the classification (for example, 985, 211), and the type of the institution (such as comprehensive, science and engineering). These data were collected from the official website through a process of web scraping. The details of this process are as follows:

```
Input : CSV file of university names ("gx.csv")
Output: A structured table that contains fields including Name,
Province, City, Address, Introduction, 985, 211, School Type,
School Attributes, as well as Featured Majors.

Read the file named "gx.csv" in order to carry out the action of
obtaining the list of universities, which is referred to as the
school_list.
For each school_name in school_list do
Construct a search URL, which is to create a URL that can be used
for searching. Specifically, the search URL is
https://gkcx.eol.cn/search?key=school_name. In this process, we
are in the situation of constructing a URL that is related to a
search function. And this URL is designed to perform a search
operation where the key parameter is set as the school name.
    Send HTTP request; parse the returned HTML page
    If university detail page link is not found then
```

DOI: 10.54254/2755-2721/2025.LD27290

```
        Record as missing; continue to next school_name
    End If
    Construct detail page URL; send HTTP request; parse HTML
    Extract fields:
        - Name
        - Province, City
        - Address
        - Introduction
        - 985/211 classification
        - School Type (e.g., Science & Engineering, Comprehensive)
- School Attributes (for example, in the situation where schools
have the attribute of being Double First - Class or being
affiliated with the Ministry of Education)
        - Featured Majors
    Store the extracted fields as a record in results
Sleep in a random manner for a period that ranges from 1 second
and then 2 seconds so as to carry out the action of avoiding the
detection of anti - crawling.
End For
Write results to output CSV file
```

The data regarding the admission scores were obtained from the official examination website of Liaoning Province. To be more specific, in the situation where one logged in to the website, the file named "2024 Liaoning Province General Undergraduate Admission Scores for Additional Rounds" was extracted. This kind of dataset contains structured information about the university, the major, the subject stream (which is either Physics or History), the batch, and the minimum admission score. These pieces of information are connected in the graph.

3.1.2. Major-interest graph

This graph manages to capture the association that exists between the categories of undergraduate majors and the interests of students. To begin with, we went through the process of gathering the complete list of undergraduate majors that are regulated by the Ministry of Education (MoE). These undergraduate majors were then categorized into 91 predefined major classes. In the process of classifying them into these classes, a semi - automatic approach was adopted. Specifically, large language models (LLMs) such as DeepSeek and ChatGPT were used for this semi - automatic classification. After that, a hobby taxonomy was generated with the help of DeepSeek. This involved mapping each interest category to the major class that is the most relevant to it. The algorithm presented below outlines the logic that is used to assign hobbies to the categories of majors:

```
Input : CSV file with fields [Major Name, Major Category]
Output: A CSV file that has fields which include Major Name, Major
Category as well as Hobbies
Function generate_hobby_templates():
Create a dictionary that performs the action of mapping each Major
Category to a list which has a quantity ranging from 5 to 10 hobby
templates.
Return the dictionary
```

DOI: 10.54254/2755-2721/2025.LD27290

```
Function read_majors_from_csv(file_path):
    Initialize empty list majors
    Open the CSV file at file_path
    For each row in CSV:
        Extract Major Name and Major Category
        Append a dictionary with these fields to majors
    Return majors

Function assign_all_hobbies(majors, hobby_templates):
    For each major in majors:
        Get class_name from major['Major Category']
        If class_name exists in hobby_templates:
            Assign hobby_templates[class_name] to major['hobbies']
        Else:
            Assign an empty list to major['hobbies']
    Return updated majors

Function save_to_csv(data, file_path):
    Open file at file_path for writing
Write the row for the header: that is to say, write a row that
contains the content of "Major Name", "Major Category" as well as
"Hobbies".
    For each item in data:
        Join hobbies into a single string separated by commas
        Write a row with major name, category, and hobbies string

Function main():
    majors = read_majors_from_csv("11.csv")
    hobby_templates = generate_hobby_templates()
    save_to_csv(majors_with_hobbies, "hobbies.csv")
    Print "Completed! Processed X majors."

main()
```

The types of nodes and the types of relationships of the integrated knowledge graph are presented in such a way that they are shown in Table 1 and also in Table 2.

Table 1. Node types in the knowledge graph

Node Type	Description
University	Higher education institution entity
Major	Specific academic discipline
MajorClass	Academic major category (e.g., Finance)
Hobby	Student interests and preferences
Province	Geographic location of university
Score	Admission score record
Stream	Exam stream: Physics or History
Batch	Admission batch (e.g., Regular)
Year	Admission year (e.g., 2024)

DOI: 10.54254/2755-2721/2025.LD27290

Table 2. Relationship types in the knowledge graph

Relationship	Description
HAS_HOBBY	Connects a major class to relevant hobbies
BELONGS_TO	Maps a specific major to a major class
ADMITTED_WITH_SCORE	Associates a major with its admission score
BELONGS_TO_STREAM	Links a score to its subject stream
BELONGS_TO_BATCH	Denotes admission batch of a score record
BELONGS_TO_YEAR	Identifies the year of a score record
OFFERS	University offers the given major
LOCATED_IN	Geographical relation between university and province
Relationship	Description
HAS_HOBBY	Connects a major class to relevant hobbies
BELONGS_TO	Maps a specific major to a major class
ADMITTED_WITH_SCORE	Associates a major with its admission score
BELONGS_TO_STREAM	Links a score to its subject stream
BELONGS_TO_BATCH	Denotes admission batch of a score record
BELONGS_TO_YEAR	Identifies the year of a score record
OFFERS	University offers the given major
LOCATED_IN	Geographical relation between university and province

The full ontology structure is illustrated in Figure 2.

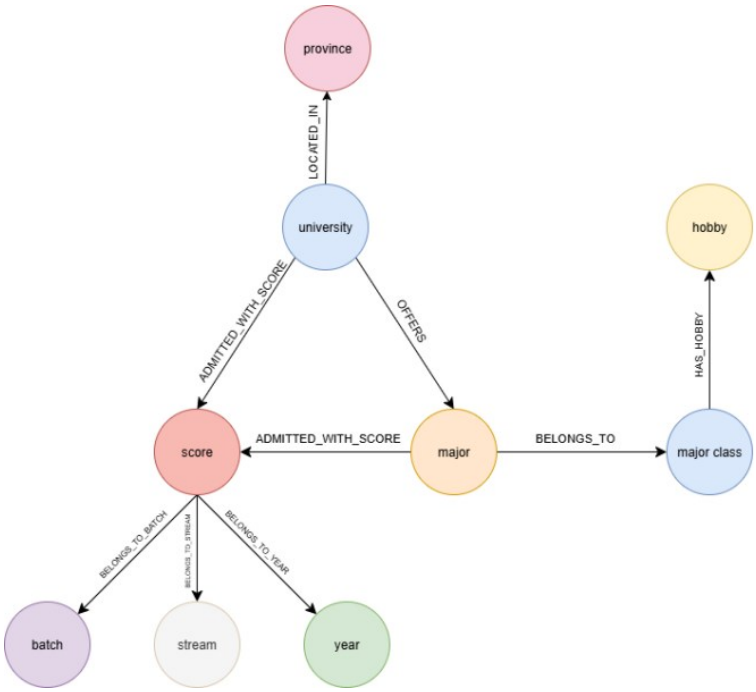


Figure 2. The ontology of the knowledge graph

3.2. Extraction-enhanced recommendation mechanism

To make the constructed knowledge graph be put into operation for real - time and user - centric recommendations, we put forward an extraction - enhanced matching framework. This framework carries out a systematic process on natural language user queries and offers ranked university - major recommendations. When a user submits a query like “I am interested in quantitative trading, scored 546, and want to study in a university in Henan. I belong to the Physics stream”, at this time, the system gets involved in the following comprehensive steps.

DOI: 10.54254/2755-2721/2025.LD27290

3.2.1. Information extraction

The framework makes use of natural language processing techniques so as to carry out an accurate parsing of the user's input. In the process of doing this, it extracts a set of key attributes. These key attributes include personal interests, examination score, preferred geographic region, as well as subject stream (for example, Physics or History), and it is not limited to just these. This kind of process guarantees that both the explicit user intents and the implicit user intents are captured and then mapped to the corresponding entities in the knowledge graph.

3.2.2. Graph-based candidate generation and filtering

The system starts the process of initiating the search. It does this by carrying out the action of mapping the user interests that have been extracted to the Hobby nodes which are present within the graph. After that, it then goes on to locate the associated MajorClass (for example, Finance) that has the best alignment with these interests and then proceeds to retrieve all of the relevant Major nodes that are under this particular category.

Next, the system takes on the task of filtering for universities. It does so by leveraging the LOCATED_IN relationship in the graph. These universities are the ones that offer these majors and are geographically located in the province that the user has specified (for instance, Henan).

For each pair of university - major that has been identified, the system further takes steps to constrain the candidate set. It does this by examining the historical admission records. More specifically, it selects the Score nodes that correspond to the user's examination stream (such as Physics), the designated batch (like regular undergraduate), and the target year (for example, 2024). Only those entries where the absolute difference between the user's score and the official admission score does not go beyond a defined threshold (say, 50 points) are kept.

The final set of eligible university - major pairs is then put through the process of being ranked. The ranking is done according to the proximity of the admission score to the user's actual score. This ranking is carried out in ascending order of |score_user - score_admission|. This ensures that the most competitive and achievable options are given the highest priority.

3.2.3. Recommendation output and interpretability

The system offers the user a ranked list of recommended universities and majors. It also provides key information, including admission scores, score differentials, and the location of the institutions. In the process of presenting this information, the system makes it transparent. This kind of transparent presentation does not only support the user to make well - informed decisions but also offers interpretability. It does so by explaining the reasons behind each recommendation.

Through the process in which the end - to - end and extraction - enhanced recommendation is carried out, the system effectively conducts an integration of user preferences, interests, as well as academic qualifications with historical admission data that is embedded in the knowledge graph. This kind of approach provides recommendations that are personalized, accurate, and interpretable. As a result, it offers support to students so that they can make optimal choices for their academic and professional futures.

4. Methodology

4.1. Knowledge Graph construction

In the course of this work, we carried out an action of leveraging the state - of - the - art GLM - 4 - plus large language model. We did this in order to systematically engage in the process of structuring

DOI: 10.54254/2755-2721/2025.LD27290

diverse entities. These entities include universities, majors, admission scores, and student interests. We then combined them into a comprehensive knowledge graph. This kind of knowledge representation provides support for the discovery of entity patterns. This enables the system to carry out the operation of matching candidate university - major combinations to individual student profiles. Moreover, our system gives users the ability to submit natural language queries. Then, in response, it can provide structured and data - driven recommendations.

To offer support for these functionalities, we carried out the process of constructing five core data files. These data files include example.xlsx, gx.csv, hobbies.csv, lishi.csv, as well as wili.csv. The university profile information was obtained through the process of web scraping from the official platform. In the meantime, the minimum admission scores for each major in 2024 were sourced from the official website of the Liaoning Provincial Education Examinations Authority.

The types as well as the quantities of the nodes and the relationships in the process of constructing the knowledge graph are summarized in Table 3.

Table 3. The types and quantities of nodes and relationships in the knowledge graph of university - major admission

Node Type	Quantity	Relationship Type	Quantity
University	2,867	HAS_HOBBY	453
Batch	14,298	BELONGS_TO	10,581
Hobby	438	ADMITTED_WITH_SCORE	28,596
Major	4,298	BELONGS_TO_STREAM	14,298
MajorClass	91	BELONGS_TO_BATCH	14,298
Province	33	BELONGS_TO_YEAR	14,298
Score	14,298	OFFERS	14,298
Stream	14,298	LOCATED_IN	2,822
Year	14,298		

In this context, we are referring to the various types and the corresponding quantities of nodes as well as relationships that exist within the knowledge graph related to the process of university - major admission. This knowledge graph is a complex structure where different elements interact, and understanding the types and quantities of nodes and relationships is crucial for a comprehensive analysis of the university - major admission system.

This kind of extensive schema makes sure that it can achieve broad coverage and fine - grained granularity. It does this by going through a process where it actively secures these aspects. In this way, it provides the basis for the process of carrying out robust matching and then providing recommendations.

4.2. Recommendation system performance evaluation

To carry out a rigorous evaluation of the effectiveness of the recommendation framework that we have proposed, we went ahead and designed a set of performance tests. We did this by making use of the example.xlsx dataset. This dataset contains eight key fields, that is to say, columns A then B then C then D then E then F then G then H. These fields are as follows: Interest, which refers to the student's stated interest; Target Province, which represents the location of the desired universities; Score, which is the student's entrance exam score; Track, which can be either the Physics or the History stream; Final Major, which is the major that the student actually chose; Final University, which is the university that the student actually attended; final_decision, which is the recommendation generated by our model; and baseline, which is the recommendation from a baseline model.

DOI: 10.54254/2755-2721/2025.LD27290

A detailed description regarding each of the input fields and output fields is presented within Table 4. In this context, a description is given in the process of going through Table 4, and it is a description that is carried out to provide a clear account of each input field and each output field.

Table 4. The input variables and the output variables for the recommendation system of college entrance examinations

Field	Description
Interest	User-specified interests
Target Province	Intended province for university selection
Score	User's entrance examination score
Track	Physics/History stream selection
Final Major	Major actually chosen by the user
Final University	University actually chosen by the user
final_decision	Recommendation provided by the proposed system
baseline	Recommendation provided by the baseline system

We carried out a process of conducting a comparative evaluation of our knowledge graph - enhanced system when it is put in a situation of being compared against a conventional baseline. This conventional baseline is one that relies only on the process of performing keyword - based matching and search. The main difference lies in the fact that our system goes through the process of applying extraction - enhanced reasoning. This reasoning is based on the situation where a structured knowledge graph is used as a foundation. On the other hand, the baseline only engages in the process of performing surface - level language matching. As a consequence, the empirical data regarding the actual final choices of universities and majors show that our system, in the process of making recommendations, significantly outperforms the baseline in the aspect of recommendation accuracy. Tables 5 and 6 present the results of the comparative process for the accuracy of major recommendation and the accuracy of university recommendation, respectively.

Table 5. Major recommendation accuracy

	Top 1	Top 3	Top 5	Top 10	Top 15
Ours	0.230769	0.538462	0.923077	1	1
Baseline	0	0	0	0.230769	0.307692

Table 6. University recommendation accuracy

	Top 1	Top 3	Top 5
Ours	0.307692	0.769231	1
Baseline	0	0	0.692308

As it is shown in the tables, in the situation where our knowledge graph - driven recommendation system is in operation, it manages to bring about substantial improvements in the process of evaluating all evaluation metrics. This system doesn't just carry out an action to enhance the accuracy of the top - ranked recommendations. Instead, as the candidate set goes through the process of expanding, it also shows a performance that is robust. This situation underscores the practical utility and the reliability of this kind of proposed approach for the purpose of providing support to the decision - making process regarding college entrance.

5. Discussion

The process of college entrance preference selection, which serves as a crucial and pivotal process that has an impact on the future development of students, frequently brings about a great deal of confusion and uncertainty for both the candidates and their families. In the situation where the existing systems for querying historical admission data offer some reference value, they usually depend on

DOI: 10.54254/2755-2721/2025.LD27290

structured input formats. As a result, they lack support for natural - language interaction, which restricts their usability and the user experience. At the same time, when the Large Language Models (LLMs) have emerged, this has created new opportunities for more natural and effective human - computer dialogue. Nevertheless, the inherent “hallucination” problem in LLMs, in other words, the potential to generate information that is inaccurate or not based on reality, gives rise to concerns about the reliability of their outputs.

In order to deal with these challenges, this research study puts forward a methodology for the construction of a knowledge graph regarding university - major admission information. This methodology is based on the situation where historical admission records are taken into account. After the construction of this knowledge graph, we then go on to design a framework for a recommendation system. This framework integrates strategies that enhance the process of extraction so as to make it easier to achieve a more precise matching between candidates and academic programs. The results of the experiments show that the proposed way of doing things leads to significant improvements in the accuracy of the recommendation.

6. Conclusion

There exist two primary limitations within the scope of this study. To begin with, the process of interest mapping is not entirely accurate. This is because the hobbies.csv file makes use of the national undergraduate major catalog that has been published by the Ministry of Education. At the same time, the lishi.xlsx and wuli.xlsx files are based on the admission data from the year 2024 provided by the Liaoning Provincial Education Examination Authority. As a consequence, the situation where there are inconsistencies in major names among different sources has led to incomplete matching. In future work, this issue could be dealt with by carrying out the expansion and standardization of the major datasets so as to achieve better coverage. Secondly, the volume of data is restricted. This is due to the fact that the current system only takes into account the admission records from Liaoning Province in the year 2024. In order to enhance the robustness and practical value of the system, future research should engage in the collection and integration of multi - year data from multiple provinces by making use of the methods that are described in this paper.

References

[1] MCheng, Y., & Hamid, M. O. (2025). The social impact that the Gaokao has in China: A critical review of the research that has been carried out. Language Testing in Asia, 15(1), 22.

[2] Chen, S., Xie, J., Wang, G., Wang, H., Cheng, H., and Huang, Y. (2024). In the process of moving from a situation where it is score - driven to a situation where value - sharing occurs: Understanding the way that Chinese families make use of AI in order to provide support for the decision - making process of college applications. arXiv preprint arXiv:2411.10280.

[3] Klimashevskaia, A., Jannach, D., Elahi, M., along with Trattner, C. (2024). A survey that is carried out on the popularity bias within recommender systems. User Modeling and User - Adapted Interaction, 34, which is the volume number, and 5, which is the issue number, and the pages range from 1777 and then to 1834.

[4] Qiao, L., Tang, Z., and Zhang, W. (2024). In the process of filling out the college entrance examination application, the dilemmas and the ways to handle the dilemmas in the major selection. Journal of Education and Teaching Research, 38(3), 52 to 67.

[5] Adel, A., and Alani, N. (2025). Is it possible for generative AI to carry out the task of synthesizing literature in a reliable manner? The process of exploring the hallucination issues that exist in ChatGPT. AI and Society, from page 1 and then to page 14.

[6] Xu, Z., Jain, S., and Kankanhalli, M. (2024). The occurrence of hallucination is something that cannot be avoided: It is an innate kind of limitation that large language models possess. arXiv preprint arXiv:2401.11817.

[7] Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., … and Ruetsch - Chelli, C. (2024). In the process of conducting systematic reviews, the hallucination rates and the reference accuracy of ChatGPT and Bard are studied. A comparative analysis is carried out. This research is published in the Journal of Medical Internet Research, with the volume of 26 and the issue of 1, specifically e53164.

DOI: 10.54254/2755-2721/2025.LD27290

[8] Wang, X., and Cheng, G. (2024). A survey on the process of extractive knowledge graph summarization: Applications, approaches, evaluation, as well as future directions. arXiv preprint arXiv:2402.12001.
[9] Yang, X., & Tan, L. (2022). The construction of an accurate recommendation model for learning resources of a knowledge graph in the situation where deep learning is exerting its influence. Scientific Programming, 2022(1), 1010122.
[10] Yan, L., Sha, L., Zhao, L., Li, Y., Martinez - Maldonado, R., Chen, G., … and Gašević, D. (2024). In the process of considering the practical and ethical challenges of large language models in the field of education, this is a systematic scoping review. The British Journal of Educational Technology, 55, with the first issue, has a page range from 90 to 112.
[11] Carver, J. C., Kendall, R. P., Squires, S. E., and Post, D. E. (In the month of May in the year 2007). Software development environments for scientific and engineering software: A series of case studies. In the process of the 29th International Conference on Software Engineering (ICSE'07) (in the range from page 550 then to page 559). IEEE.
[12] Li, Y., Wang, S., Ding, H., and Chen, H. (In the month of November in the year 2023). Large language models within the field of finance: A survey. In the process of the fourth ACM international conference on AI in finance taking place, it was presented in the Proceedings of this event (in the pages that range from 374 to 382).
[13] Zhang, Y., Chen, X., Jin, B., Wang, S., Ji, S., Wang, W., and Han, J. (2024). A comprehensive survey that conducts an in - depth exploration of scientific large language models and their applications in the process of scientific discovery. arXiv preprint arXiv:2406.10833.
[14] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … and Liang, P. (2021). In the situation where one is considering the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
[15] Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., and Wang, P. (In the month of April in the year 2020). K-bert: The process of enabling language representation by means of a knowledge graph. In the process of the Proceedings of the AAAI conference on artificial intelligence (In the volume of 34, and the number of 03, and the pages from 2901 to 2908).