A Study on Religious Beliefs among College Students Based on FP-Growth Algorithm

Weihua Yao [a], Shunfei Yan [b*], Yanshuang Wang [c]

Xi'an University of Technology School of Economics and Management Xi'an, China

[a]645903196@qq.com, [b*]1154620052@qq.com, [c]706495200@qq.com

Abstract. In the backdrop of ongoing societal development, the trend towards secularization of religion becomes increasingly apparent. The infiltration of religious thoughts and influences within university campuses is progressively on the rise. In this context, countering and preventing the incursion of religious influences within educational institutions has emerged as a crucial task in the domain of ideological and political education for college students. This paper employs the FP-Growth algorithm to analyze the religious beliefs among college students, aiming to explore the influences of factors such as background, social environment, and family upbringing on the psychological changes related to religious faith among this demographic. Through the analysis of college students' religious beliefs, it unveils the psychological changes and general patterns associated with their embrace of religion, contributing to a better understanding of the motivations and transformations behind their religious beliefs. Ultimately, this study presents a set of strategic recommendations for universities to resist and prevent the infiltration of religious influences, guiding college students towards upholding a steadfast Marxist faith and cultivating a proper worldview, philosophy of life, and value system.

Keywords: FP-Growth Algorithm; Religious Beliefs Psychology; Religious Infiltration; College Students

1 Introduction

With the development and progress of society, religious beliefs are playing an increasingly important role in people's lives. In university campuses, the issue of religious beliefs among college students has always been a topic of concern. Some college students, upon exposure to religion, develop thoughts of converting to a religious faith. This phenomenon has attracted the attention of many experts and scholars [1]. The reasons for college students adopting religious beliefs and their psychological changes are complex issues that require in-depth research and analysis [2].

In this context, this article proposes a psychological analysis method for college students' religious beliefs based on the FP-growth algorithm. The FP-growth algorithm is an efficient data mining technique used to discover frequent itemsets within data, thus revealing patterns and trends present in the data [3]. This article applies the algorithm to

The original version of this chapter has been revised. A correction to this chapter can be found at https://doi.org/10.2991/978-94-6463-276-7_55

© The Author(s) 2023, corrected publication 2024

X. Ding et al. (eds.), *Proceedings of the 2023 4th International Conference on Big Data and Social Sciences (ICBDSS 2023)*, Atlantis Highlights in Social Sciences, Education and Humanities 12,
https://doi.org/10.2991/978-94-6463-276-7_29

the analysis of college students' religious beliefs, aiming to uncover patterns and trends in the psychological changes related to religious beliefs among college students. The goal is to provide reference and recommendations for higher education institutions to develop strategies against and prevent religious infiltration.

The study in this article unfolds through several aspects: firstly, introducing the current status and reasons for college students' religious beliefs, analyzing the psychological changes and influencing factors of such beliefs. Secondly, explaining the principles and applications of the FP-growth algorithm, analyzing its advantages and limitations in the analysis of college students' religious beliefs [4]. Then, through the analysis of experimental data, exploring the patterns and trends in the psychological changes of college students' religious beliefs, revealing the psychological mechanisms and needs behind such beliefs. Lastly, combining the experimental results, presenting strategies and suggestions to counter and prevent religious infiltration, thereby offering valuable guidance and reference for higher education institutions.

The significance of this research lies in the fact that through the psychological analysis of college students' religious beliefs based on the FP-growth algorithm, it can uncover patterns and trends in their psychological changes related to religious beliefs. This provides a reference for developing strategies against and preventing religious infiltration in higher education institutions [5]. Additionally, the results of this research can also contribute to the psychological education of college students, helping them approach religious beliefs rationally, and enhancing their mental resilience and intellectual awareness.

2 FP-growth algorithm

Association rule mining is to mine useful, potential, and interrelated knowledge between data items from massive data, and the association rule mining analysis between courses is to mine the interrelationships between courses. Typical association rule mining algorithms include Apriori algorithm and FP-growth algorithm. The FP-growth algorithm is an association rule analysis and mining algorithm proposed by Han Jiawei et al. in 2000. Compared with the Apriori algorithm, the FP-growth algorithm does not generate candidate item sets, and the FP-growth algorithm is more efficient than the Apriori algorithm [6]. FP-growth algorithm is currently widely used in various fields. Therefore, this paper uses the FP-growth algorithm to mine the correlation between courses.

2.1 Related definition

The support degree support indicates the proportion of a certain item or a certain rule that appears in all its historical transaction data or the proportion of all items contained in a certain rule that appear in all transaction histories at the same time, which can be expressed as the following formula:

$$support(A => B) = support(A \cup B)/P(AB) \quad (1)$$

The confidence of a rule measures the proportion of successors that occur when the predecessor appears in the transaction. Confidence reflects how reliable the rule is. The confidence calculation of the association rule A=>B is shown in the following formula:

$$confidence(A => B) = support(A \cup B)/support(A) \tag{2}$$

The minimum support is represented by min-sup, and the minimum confidence is represented by min-conf. The thresholds of these two values can be set according to their own needs. If the support of A=>B is greater than the given min-sup threshold and the confidence of A=>B is greater than min-conf, then A=>B is established, that is, rule A can be derived from rule B.

2.2 Detailed process of FP-growth algorithm

The main content of the FP-growth algorithm can be divided into two core steps. The first is to compress the database into the FP-tree tree (FP-tree is a special prefix tree composed of frequent item header table and item prefix tree), and then mine frequent item sets in the FP-tree tree. The flow of FP-growth algorithm is shown in Figure 1.

The detailed steps of the FP-growth algorithm are as follows:

The first step is to construct FP-tree. Scan the database and calculate the support of each item in the database. If the support of the item is greater than a given threshold, the item is a frequent itemset, and the frequent itemsets are stored in the database in order of support. Scan the database for the second time, read the database in turn and save it in the FP-tree according to the path, repeat this step until the database is read, that is, the construction of the FP-tree is completed [7].

The second step is to mine frequent itemsets in the FP-tree. After the FP-tree tree is constructed, traverse the FP-tree in the order from the leaf node to the root node in the FP-tree, and create a conditional pattern base for each node in order (the conditional pattern base is based on the searched element item as The path set at the end represents all the content between the element item you are looking for and the root node of the tree), and then create a conditional pattern tree based on the created conditional pattern base, and then mine frequent patterns in the conditional pattern tree, and then dig out Frequent itemsets, all frequent itemsets can be obtained at this time[8].

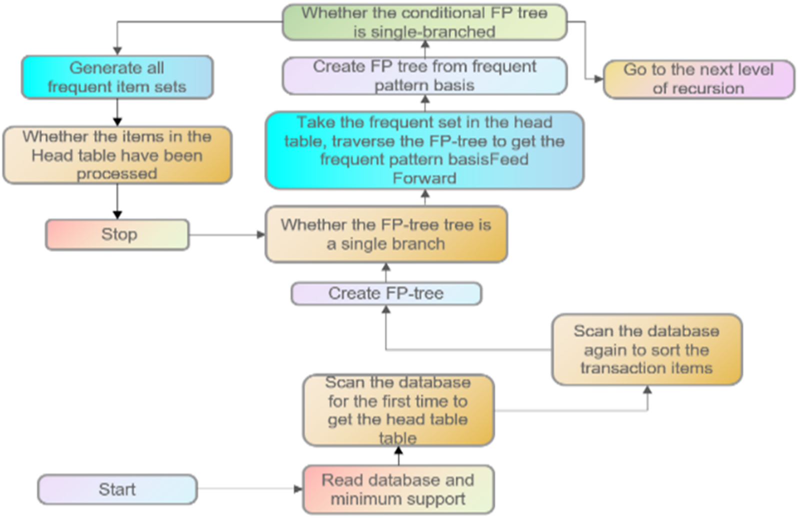

Fig. 1. Flow chart of FP-growth algorithm

2.3 The process of building FP tree

Table 1. Partial data element abstract table

Transaction ID	Element items in transactions
01	l,q,h,j,y
02	q,p,o,w,v,u,n,m
03	q
04	l,o,m
05	p,l,o,q,x,n,y
06	p,q,o,x,m,n

Taking the data in Table 1 as an example, first, scan all data and count the number of occurrences of a single data item. Based on the set minimum support, remove data items that are less than the minimum support. These removed data items will not participate in the subsequent FP tree construction. Because according to the Apriori principle, if a data item or itemset is infrequent, then the superset containing that data item or itemset is also infrequent. Scan Table 1 and count the number of occurrences of each data item. l appears 3 times, m appears 3 times, n appears 3 times, o appears 4 times, p appears 3 times, q appears 5 times, x appears 2 times, y appears 2 times, and h, j, w, v, u only appear once. We set the minimum support to 2, so the number of occurrences of x, y, h, j, w, v, and u is less than the minimum support, so we delete x, y, h, j, w, v, and u, and they are not included in the construction of FP tree. We place the data items that meet the minimum support level in the head Table and sort them from high to low according to the support threshold of each data item, as shown in Figure 2.

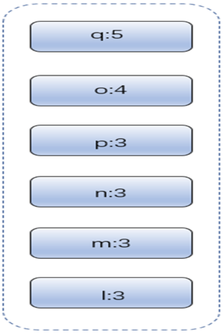

Fig. 2. Head table corresponding to Table 1

The first scan of the dataset is to count the number of occurrences of each data item and filter out infrequent data items based on the set minimum support level. The second scan of the dataset allows for only frequent data items without considering those that are not frequent. Because when constructing FP tree, the same set of items is in the same path of FP tree, and the tree branches only when the subsequent element items are different, the size of FP tree is smaller than the size of the original dataset. The most ideal scenario is that FP tree has only one path, and all transactions are included in this path. This is an idealized situation, and our actual dataset is complex and diverse, which basically does not happen. On the contrary, when FP tree has multiple paths, that is, the itemset of each transaction is unique, the size of such FP tree is the same as the size of the original dataset, and this situation usually does not occur because the data is interrelated and interdependent. When building an FP tree, scan each transaction first to find the frequent items in each transaction. If the found frequent items have a node with a corresponding path in the FP tree, modify the number of occurrences of the node directly. Otherwise, create a new node and add a pointer to the node in the head Table. If two identical nodes have different paths, a link will also be created between the two nodes.

Taking the data in Table 1 as an example, the process of constructing FP tree is as follows:

When scanning transaction 01, h, j, and y are all infrequent data items, so there is no need to consider that transaction 01 is equivalent to {l, q}. After scanning transaction 01 for FP tree, the node information of q is 1, indicating that q appears once, and l appears once on the path prefixed with q. Two links will be created in the head Table, where q on the head Table points to the q node in FP tree, and l on the head Table points to the l node in FP tree, which facilitates searching between similar element items. When scanning transaction 02, remove the infrequent items, and transaction 02 becomes {q, o, p, n, m}. The information of node q becomes 2, connecting node o, node p, node n, and node m. These four nodes are all connected after node q for the first time, so the node information is all 1. Head Table creates a link between similar nodes and FP tree; When scanning transaction 03, remove infrequent items, and transaction 03 becomes {q}, so the information of node q becomes 3; When scanning transaction 04, remove infrequent items, and transaction 04 becomes {l, o, m}. There is no q data item in transaction 04, so FP tree forks out a new path, and finally links between similar nodes; When scanning transaction 05, remove the infrequent items, and transaction 05 becomes {p, l, o, q, n}. The transaction contains q data items, so it is also prefixed with the node q. The node information of q, o, p, and n is added by 1, respectively. Node m is different from node n, so the final node m forks; When scanning transaction 06, remove the infrequent items, and transaction 06 becomes {p, q, o, m, n}. There is an identical path in FP tree, so the node information on this path is added by 1. All transactions have been scanned, and the final FP tree is shown in Figure 3

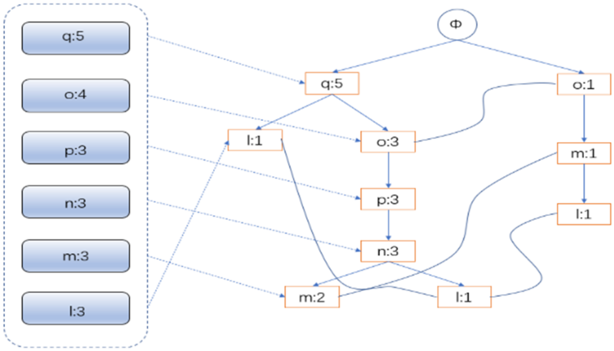

Fig. 3. FP-tree diagram

3 Correlation Analysis of College Students' Religion Factors Based on FP-growth Algorithm

3.1 Data selection

The data used for correlation analysis in this paper is collected by conducting online questionnaire surveys and offline questionnaires for college students in the school. In the end, 262 online questionnaires were effectively recovered, and 200 offline questionnaires were recovered. The objects of the survey are the students of our school, 121 males and 141 females filled in the online questionnaires; 100 females and 100 males filled the offline questionnaires. This survey has received strong support from alumni, so this survey has certain value and certain credibility.

3.2 Data preprocessing

Data preprocessing is a very important link in data mining. Data preprocessing can improve the accuracy and effectiveness of association rule mining. Before association rule mining, it is necessary to extract data from the data and then generate the required data from the data format.

(1) Data extraction

In the data obtained by the test, since the factors such as name and test date have no influence on the mining results, these factors can be removed.

(2) Data cleaning

There are still a lot of dirty data in the data after data extraction. It is difficult to build a good mining model without data cleaning, so the data needs to be cleaned. Data cleaning includes missing value processing, noise data processing, abnormal data processing, duplicate data checking, and data validity verification [9]. After preprocessing the questionnaire data of college students' religious beliefs, they are entered into Excel, and then association rules are mined.

3.3 Association Rules Mining

The FP-growth algorithm is applied to the psychological analysis of college students' religious beliefs, and the implicit correlation between factors is excavated. Table 2 is the process of calling FP-growth algorithm mining.

Table 2. The process of calling FP-growth algorithm

```
import pandas as pd
from mlxtend.frequent_patterns import fpgrowth
from mlxtend.frequent_patterns import association_rules
# read excel file
excel_file = "data.xlsx"
data = pd.read_excel(excel_file)
# One-hot encode data to convert discrete features to 0/1 values
data_encoded = pd.get_dummies(data)
# Frequent Itemset Mining Using FP-growth Algorithm
frequent_itemsets = fpgrowth(data_encoded, min_support=0.1, use_colnames=True)
# Generate association rules based on frequent itemset mining results
association_rules_result = association_rules(frequent_itemsets, metric="lift", min_thresh-
old=1.0)
# Filter out meaningful association rules based on support and confidence
significant_rules= association_rules_result[(association_rules_result['support'] >= 0.1) &
(association_rules_result['confidence'] >= 0.5)]
# print association rules
print(significant_rules)
```

3.4 Association Rules Mining

Based on the FP-growth algorithm, we analyzed the belief questionnaires filled out by college students. The experiment sets the minimum support degree as 0.1 and the minimum confidence degree as 0.6. Table 3 lists some results of the program operation.

Table 3. Partial mining results

Rule	Confidence	Support
Personal growth and self-knowledge =>Thinking about life, death and existence	0.6751	0.1245
Culture and Traditions => Religious Leaders and Mentors	0.7581	0.2365

Seeking Ethics =>Personal Dilemmas and Challenges	0.8345	0.3124
Family Background =>Finding Meaning and Purpose	0.6313	0.1123
Knowledge and Education=>Science and Rational Thinking	0.7234	0.2214
Social change and value transformation =>Social pressure and social identity	0.6898	0.2122

Taking the rules {personal growth and self-knowledge}→{thinking about life, death and existence}, {culture and traditions}→{religious leaders and mentors} as examples, the results of mining can be explained as follows: exploring self-knowledge in the process of growing up of students were 67% likely to ponder life, death, and existence, as religious beliefs may provide a framework and answers about these questions, leading to religious beliefs. In some cultures, religious beliefs are part of the culture and traditions, and there is a 75% probability that students who have religious beliefs and cultural traditions will lead to religious beliefs through encounters with religious leaders and mentors.

4 Discussion and Conclusion

4.1 Results and Findings

By using the FP growth algorithm to analyze the religious factors of college students, we obtained the following results and findings: Firstly, we found that approximately 0.35% of college students in our sample believe in religion. This indicates that religion still has a certain influence on university campuses [10]. Secondly, we found that the main factors affecting college students' religious beliefs are their family background and inheritance of religious beliefs. This is consistent with previous research results. In addition, we also found that the degree of religious adherence among college students is related to factors such as gender, major, and grade. For example, female college students are more likely to believe in religion, students majoring in social sciences are more likely to believe in religion, and senior students are more likely to believe in religion. Finally, we also analyzed the correlation between factors of religious belief and found that students' religious belief is not achieved overnight, but rather is related to factors. For example, taking the relationship between {culture and tradition} and {religious leaders and mentors} as examples, in some cultures, religious belief is a part of culture and tradition, Students with religious beliefs and cultural traditions have a 75% chance of leading to religious beliefs through contact with religious leaders and mentors. We can take targeted measures by clarifying the weight of these factors on college students' religious beliefs, in order to use big data to empower and more conveniently help ideological and political education teachers in universities discover and solve problems in a timely manner.

4.2 Results and Findings

The results and findings of this study have certain significance and contributions to religious management and education on university campuses. Firstly, this study provides a new method of using the FP growth algorithm to analyze the religious factors of college students. This method can help university administrators and educators better understand the situation and reasons for college students' religious beliefs, and provide reference for formulating more effective religious management and education policies. Secondly, the results of this study can help university administrators and educators better understand the characteristics and laws of university students' religious beliefs, thereby better guiding and educating university students to cultivate correct worldviews, life philosophies, and values.Finally, by studying the interrelationships and internal correlations between religious factors, we can conduct in-depth analysis and weight assessment of these factors to understand their impact on college students' beliefs, and then develop targeted measures. At the same time, using big data analysis technology, we can more efficiently help ideological and political education teachers in universities identify and solve problems in a timely manner, and improve the quality and efficiency of education. The results of this study can provide reference and inspiration for research in related fields, and provide a foundation and direction for future research.

4.3 Future Prospects

There are also some shortcomings in this study, and future research can be carried out from the following aspects: Expand the sample range and coverage, and verify the universality and representativeness of the results[11]; Exploring the application and comparison of different algorithms and methods in the study of religious factors among college students; Study the impact of religious beliefs on the mental health and academic performance of college students, as well as how to carry out effective religious management and education[12]; Explore the relationship between college students' religious beliefs and social, cultural, historical and other factors, and gain a deeper understanding of the background and reasons for college students' religious beliefs

Acknowledgment

Xi'an University of Technology Humanities and Social Science Research Coun-selor Backbone Special Project "Research on the Construction of the Guarantee System for the Growth of Post-00 Universities" (2018FZ018); Xi'an University of Technology "all-round education" Special Project "'all-round education' Com-prehensive Reform Demonstration College Cultivation Project" (XYR2101)

Reference

1. Wang, L.F. (2016) Introduction to Marxist Religious View. Zhejiang University Press, Hangzhou. https://book.qq.com/book-detail/22927471
2. Wang, L.B. (2018) Textual Interpretation of "Religion is the Opium of the People" from the Perspective of Marx's Early Thought. Science and Atheism, 116:22-27.

https://kns.cnki.net/kcms2/article/abstract?v=jJlylHAC3fqoBKbwdmY623HVHMGNqsTZ-vbMkPleUAWcUR-WZs7DrABJ1Kkv5IadeE0Gl8HTxNdZv54upWTmz6HiD0oDohVOyqNoFWCysvMP_-MoFDjV0qz8XTaNWs5ZTHGF-xKUEw04=&uniplatform=NZKPT&language=CHS

3. Shawkat M , Badawi M , El-Ghamrawy S ,et al. (2022) An optimized FP-growth algorithm for discovery of association rules [J]. Journal of supercomputing, DOI: 10.1007/s11227-021-04066-y. https://www.nstl.gov.cn/paper_detail.html?id=6d0d1fbcc84197df13fa71d08f0db587
4. Hu, J.L., Hu, W.W. (2017) Research on the countermeasures of college students' belief education based on the concept of "supply side" -- Taking college students in Jiaxing City as an example. Journal of Jiaxing University, 29:116-119. https://kns.cnki.net/kcms2/article/abstract?v=jJlylHAC3fqP5Q_kdm6bJ6hMUvg-BXFbSu7BvuqiGXCxWhI3DZga7h13ZXDO9xCMhwPELX6i_R6FSZNLh0Y4Zl7MCo8NFpXJ-vPbdl0YYPxbb3BPHqykKjznGQf0WoadlcHKw7JcDBI=&uniplatform=NZKPT&language=CHS
5. Zhang, R.H. (2016) The Perspective and Educational Approach of Belief Research on College Students. Contemporary Youth Studies, 341:50-56. https://kns.cnki.net/kcms2/article/abstract?v=jJlylHAC3fqzN1ZEzkhZTDYfVH5LuyNomi34EwTEXvVTIPAO2OB-fnIdmcTC51MlRO6Evw5B-fzv63orhyNeYGfE3snJ5_w-uEwr3es-OW_ta5fMRYjpDl5KqX9h3G76PyZLHS5xIFZo=&uniplatform=NZKPT&language=CHS
6. Yang, J. , Gong, L. , Liu, K. , & Xiu, L. . (2022). Operation anomaly monitoring of customer service data analysis platform based on improved fp-growth algorithm. Journal of Physics: Conference Series, 2209(1), 012030-. https://www.socolar.com/Article/Index?aid=100093155443&jid=100000008353
7. Chunduri, R. K. , & Cherukuri, A. K. . (2021). Scalable algorithm for generation of attribute implication base using fp-growth and spark. Soft Computing(3). https://www.nstl.gov.cn/paper_detail.html?id=142f56a83cc669d597d2b084db3dc245
8. LI, Z.Y. (2007) Analysis of Ideological Tendency and Causes of Religious College Students. Contemporary World and Socialism, 65: 120-124. https://kns.cnki.net/kcms2/article/abstract?v=jJlylHAC3fpTmTsktos29qDRCok_FFrkTGkg5pgUYiqDJxgBoZVvgUuOlp5jlWDRdVykOGQslpq6AREQePi6u2iyk1mNKeuWucSIsNbmYgMw1No-H0dMiBc2KLepwm7j&uniplatform=NZKPT&language=CHS
9. Akbar, M.. (2022). Fp-growth approach for document clustering. https://doc.taixueshu.com/foreign/rgArti0000130623370.html
10. Rani, L. N. , Defit, S. , & Muhammad, L. J. . (2021). Determination of student subjects in higher education using hybrid data mining method with the k-means algorithm and fp growth. International Journal of Artificial Intelligence Research(1). http://www.xueshufan.com/publication/3169736476
11. Xu, F. , & Lu, H. . (2017). The application of fp-growth algorithm based on distributed intelligence in wisdom medical treatment. International Journal of Pattern Recognition & Artificial Intelligence, 31(4), 232-237. https://www.nstl.gov.cn/paper_detail.html?id=fa69f608122f41c6431a73b7f7b25c39
12. Nan, L. U. , Zhe, W. , & Chun, Z. . (2003). The effect of fp-growth algorithm based on fp-tree frequent set patterns on association rule mining. Acta Scientiarium Naturalium Universitatis Jilinensis. http://en.cnki.com.cn/Article_en/CJFDTOTAL-JLDX200302015.htm

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.