ISSN 2229-5518

Graph Based Model for Recommending Financial Products

Olakunle Temitope, Awodele Oludele, Adekunle Yinka, Eze Monday

Abstract— Recommendation techniques have gradually been integrated into e-commerce systems, helping clients find certain products that better match their needs. Unfortunately, there has been no attempt to incorporate such methods into the banking domain. Many Recommendation Systems (RS) are based on subjective assessment and user ranking, resulting in inaccurate recommendations for banking institutions in particular. It has culminated in the reluctance of consumers to subscribe to bank products due to the heavy criteria and its inaccessibility. Therefore, this study established a graph-based recommender framework for the banking domain that recommends customized products and services from the experience and historical transactions of the observed customers. The developed model has helped evaluate consumer behaviour, proposing appropriate banking products to customers in a way that removes prejudice and the subjective complexity of many RSs decision-making process.

Index Terms— Banking domain, Financial products, Graph database, K-Nearest Neighbor, Decision Tree, Recommender Framework, Recommender Systems.

1 Introduction

THE eBusiness world has also advanced due to this new way of personalization which in turn has improved the financial sector [1]. The wealth of information available increases nowadays due to a continuous increase in digital information, resources and online contents [2], thereby resulting in the integration of RS in avoiding information overload. Recommender system (RS) is perceived as one of the most promising web personalization tools that assists users by providing suitable recommendations based on their preferences after extracting useful information from a large corpus of available information [3]. RS has improved over the years in terms of sophistication and connectivity to other systems due to the improved quality and quantity of available data to the RS [4],[2]. The building of a recommender system can be challenging due to quality of data, priority of recommender system goals i.e. accuracy, precision, recall novelty of suggestions, increase in size, complex problem domain and more diverse application across various domains [5]. This demand has called for development of more distinctive and advanced filtering techniques.

The unprecedented growth of competition in the financial institution has raised the importance of retaining current customers and acquires new customers. However, these can be achieved by regularly showing customers available benefits, products and services that could be enjoyed. Financial products and services can be said to be instruments that help bank customers save, invest, get insurance or get a mortgage, and become financially more empowered [6]. Due to the rapid increase in the financial data and proliferation of technology in the financial domain, the graph database system is fast outpacing other data models in terms of growth [7]. The graph data model differs from the conventional relational databases specializing in the management of interactions between such a wide range of information points, allowing the graph system developer or information researcher to adequately handle their information. With graph databases, there are no restrictions by semantically-limited information models and expensive, unexpected methods to run queries by joining the way you are forced to do the relational strategy. This advantage amidst other existing ones is very vital in analysing massive volume of real time financial data. This article presents a model for recommending financial products and services. Subsequent section covers existing works of literature. Thereafter, methodology used to develop the model is presented and the conlusion to the study follows.

2 Review of Literature

[9] propose a constraint diagnosis and repairing technique. Related to online banking and multi-domain solutions, the products are basically heterogeneous. The churn rate depends on the type of items accessed by these systems; however, we consider it low in banking environment. As these solutions offer interactive user interfaces, the interactions are explicit. It can be argued that the user preference is unstable, because it strongly depends on the actual goal of the user.

Significant work is published by [12] who formulate an instance-based credit risk assessment model for evaluating risk and return of each individual loan. This model was developed as a result of the problem associated with the effective allocation of money across different loans by accurately assessing the credit risk of each loan. The author also noted that traditional rating-based assessment models cannot meet the needs of individual investors in P2P lending, since they do not provide an explicit mechanism for asset allocation. A data-driven investment decision-making instance-based credit risk assessment model for evaluating risk and return of each individual loan was developed.

[10] looked into the complex task of recommending financial investment strategies. They proposed a framework for recommendation of asset allocation strategies which combines case-based reasoning with a novel diversification strategy to support financial advisors in the task of proposing diverse and personalized investment portfolios. They used a combination of Basic Ranking, Greedy Diversification and Financial Confidence Value (FCV) techniques. The performance of the

framework has been evaluated by means of an experimental session conducted against 1172 real users, and result showed the yield obtained by recommended portfolios overcame that of portfolios proposed by human advisors in most experimental settings while meeting the preferred risk profile. Furthermore, the diversification strategy showed promising results in terms of both diversity and average yield.

[9] present two general-purpose knowledge based recommender systems with intelligent user interface, which can be flexibly applied on various financial products. The authors prefer knowledge-based algorithms over the conventional collaborative- and content-based filtering, because they can be applied more efficiently in multi-criteria-based financial decisions. For those cases, when no results can be shown for a multi-constraint setting. [8] showed how Knowledge-based recommenders can be developed in a Human Computation based knowledge acquisition environment (PEOPLEVIEWS) and how the resulting recommendation knowledge can be exploited in a competition-based e-Learning environment (STUDYBATTLE). The PEOPLEVIEWS environment supports two basic modes of interaction - modelling mode and recommendation mode

A real-time cloud- and web-based application was developed by [11], which recommends health insurance policies. The system applies multi-attribute utility-based theory that finds the most similar products to the preference of the user based various criteria (e.g. premium, co-pay, co-insurance, benefits). Also, [6] did a research on creating a web-based application that would grant expert advice on bank products to current and prospective customers.

Since the FinTech Domain have been leveraging on Machine Learning to provide context-based services recommendation, [13] explored the usefulness of the application of Recommender Systems in the financial domain by investigating a dataset of purchases of a large set of investment assets by 200k investors. The three algorithms compared are: Bayesian Personalized Ranking algorithm, Alternating Least Squares algorithm and Word2Vec algorithm. Although the research is ongoing, evidence indicates that statistically filtered investors' decisions could be used to cluster assets: a promising starting point to build a statistically guided algorithm for recommendations.

3 METHODOLOGY

The dataset used in this research are solely from the banking sector which were retrieved from bank customers. The banking sector is a component of the finance domain and offers products and services to its users. The products include loans, mutual funds, treasury bills, asset management amongst others while services include personal banking, internet banking, and USSD banking to mention a few. In conformance with privacy laws, the banking identifiers such as account number, name were removed and numbers were used. To prevent misleading results, the dataset was cleaned. Before running the analysis, the first step taken is to ensure the proper representation and quality of data. Data processing is simply the conversion of raw data to meaningful information through a process. Data is manipulated to produce results that lead to a resolution of a problem or improvement of an existing situation. Data pre-processing was done because some machine learning algorithms require the data to be in a specific form, other algorithm can perform better if the data is prepared in a specific way; some of the raw data were not be in the best format to best expose the underlying structure and relationships to the predicted variables. The representation of the data was done by the bank since the dataset is in spreadsheet format – excel. The tasks involved in the pre-processing were; cleaning, data transformation, data reduction.

Financial statements of customers were used as the source of data. However, basic identifiers such as the account name and account number were removed. Random set of account numbers (dummy account numbers) were used just for identification purposes (i.e. a way to identify group of transactions performed by an individual). This was done in in order not to violate ethical standards and protect privacy of bank customers. Data gathered were not limited to just one financial institution. Data collection was a bit challenging as financial institutions were reluctant to provide this information. In overcoming this challenge, many friends were contacted and requested to provide financial statements for some of the banks they use. Using account number as identifiers, customers were grouped into clusters with some characteristics such as transaction frequency which was obtained from transaction date, regular inflow and average outflow of funds which was obtained from the Dr and Cr flags. The cluster was also designed to be based on products they have previously consumed. Dimensionality reduction was done to reduce the data set by identifying outliers. Using the cluster, a customer belongs to, the system looks up the business rule and makes recommendation to such customer. It is important to state that the system also uses the products consumed by customer to improve decision to be taken while providing recommendations. The data acquired from respondents were in pdf format while those from the internet were in comma separated value (CSV) format. Using an online pdf to xls converter, the pdf documents were converted and the identifiers were replaced with dummy data. The xls file were further converted into csv and these enabled data merging and a single file for all the acquired data. missing values were searched for manually and filled using expert judgement. The database approach that was used is the graph-oriented database management system. Unlike other databases, relationships take first priority in graph databases. This means that applications do not have to infer data connections using foreign keys as done in relational databases. The data model for a graph database is also significantly simpler and more expressive than those of relational or other NoSQL databases [14].

3.1 Analyzing the Spending behavior of Customers

Spending behaviors of customers were design to be grouped in clusters. This is to identify customers with similar spending behavior and financial capability as the financial ability of a customer because the spending behavior of a customer is very dependent on his financial ability. The results produced are based on the assumptions that the customer behavior follows patterns similar to past pattern and repeats in the future. This grouping approach aids decision making for the financial institutions. The decisions to be made include which target groups of customers will be encouraged to use more products, what will be the estimated probability of acceptance of new products, promotion of new products to target groups of customers, how to manage groups of customers to reach the customer satisfaction and targeted marketing. However, quality and volume of data affects the analysis of customer behavior. For this research, the account statements of the customers were used to analyze their spending behavior. From available datasets, the customers were divided into four clusters.

3.2 Loading pre processed data into the database

The preprocessed data were loaded into the graph database and modeled to have five nodes. Unlike other types of databases, the database approach is a graph-oriented database. A graph database is essentially a collection of nodes and edges. Each node represents an entity (such as a person or business) and each edge represents a connection or relationship between two nodes. Every node in a graph database is defined by a unique identifier, a set of outgoing edges and/or incoming edges and a set of properties expressed as key/value pairs. The nodes; transaction node houses data and information about transactions carried out on the account as stated in the account statement. Debit and credit transactions are identified and marked. The bank node and account nodes related directly with the transaction node. The bank node houses list of available banks. This approach was adopted for easy identification of bank and transactions carried out and products to be recommended. The bank node relates directly with the products -which are offered by the banks to customers- and the account node which is the channel of transaction between the customer and the bank. The bank node also interacts with the transaction node. The Product node holds the available bank products to be recommended to customers. This node interacts directly with the benefit node which houses possible benefits bank and customers can obtain when recommended products are used. The higher the acceptance rate of products recommended, the higher the benefit. After loading the data -using the set of queries listed in section 3.6- into the database model, the result is what is depicted in Figure 1 and Figure 2 below.

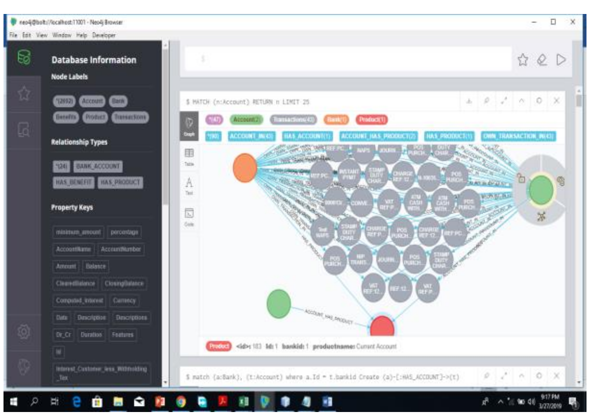

Figure 1: View of the Neo4J with the loaded data and object definitions

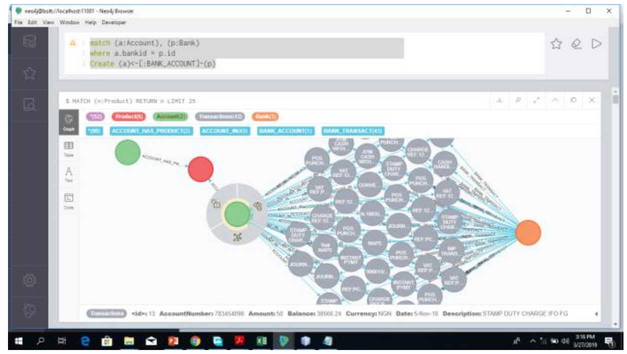

Figure 2: View of the neo4j database after data had been loaded.

As shown in figure 3 below, the relationship between the various nodes in the database were also defined.

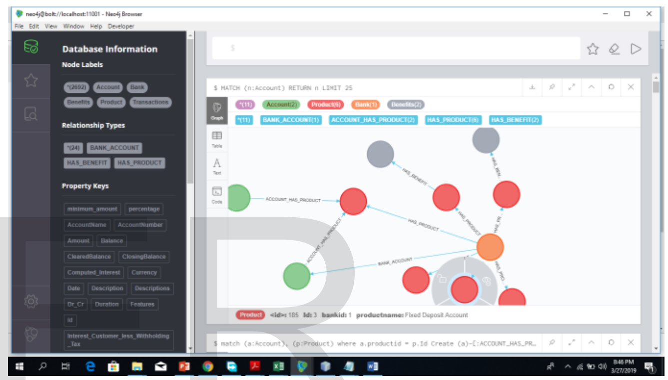

Figure 3: view of relationship between database objects.
The summary of the evaluated algorithms are presented in table 1.

Table 1: Evaluation results fo the algorithms.

Algorithm	Accuracy	Precision	recall	F measure
K-NN	99.38	0.994	0.994	0.994
C4.5	89.8	0.901	0.902	0.902
Random tree	53.62	0.538	0.538	0.538
Naïve Bayes	95.14	0.956	0.951	0.951
J48	62.77	0.631	0.628	0.628
Decision Table	99.92	0.999	0.999	0.999
BayesNet	93.44	0.943	0.934	0.934
ZeroR	62.75	0.619	0.626	0.626

Considering the performance of the algorithms, the K-NN and the decision table were selected to consist of the hybrid algorithm implemented in the system.

4 Conclusion

Recommender system gives a piece of advice about the products, information or services that the user want to know. It is an intelligent application to assist the user in a decision-making process where they want to choose one item amongst the potentially overwhelming set of alternative products or services. Unlike e-commerce sites and applications where recommender systems are mostly used, the research community had extended the use of recommender systems into diverse spheres of life, financial institution inclusive. This research had been conducted to develop an efficacious approach that will be suitable for recommending financial products and services to customers. Contrasting to the popularly used K means and fuzzy based algorithms, this research has investigated the suitability of a hybrid approach by combining the content based approach and the rule based approach in developing a recommender system as a result of the randomness observed in financial inflow and outflow of funds. It was observed that rule-based algorithms are difficult to train but have a fast recall. Unlike other financial recommendation system that tends to associate demographic information of customer so as to aid the grouping of customers, the system developed intelligently recommends products and services to customer after analyzing the financial history of the customer. An implicit feedback mechanism is also included to further improve performance of the system as the data grows. The growth of financial data had also been a disadvantage for recommender systems since they tend to possess some form of latency as the data grows significantly. This is due to the dependency of the dependencies needed to use the database approach. With the graph technology proposed in this research, the dependencies would be eliminated and the significant growth in size would have a miniscule effect on the recommendation other than to improve the performance as a result of data availability.

References

[1] Vico, D., Huecas, G., & Rodr´ıguez, J. (2012). Generating Context-aware recommendations using Banking data in a mobile recommender system. The Sixth International Conference on Digital Society. ISBN: 978-1-61208-176-2. 73-78

[2] Palomares, I., & Kovalchuk, S., (2017). Multi-View Data approaches in Recommender Systems: an overview. International Young Scientists Conference in HPC and Simulation, YSC 2017, Kotka, Finland.

[3] Choudhary, P., Kant, V., & Dwivedi, P. (2017). Handling Natural Noise in Multi Criteria Recommender System utilizing effective similarity measure and Particle Swarm Optimization. Procedia Computer Science, 115(2), 853–862. https://doi.org/10.1016/j.procs.2017.09.16

[4] Guo, G., Zhang, J., & Yorke-Smith, N. (2016). A novel recommendation model regularized with user trust and item ratings. IEEE Transactions on Knowledge and Data Engineering 28, 1607–1620.

[5] Bouzekri, E., Canny, A., Fayollas, C., Martinie, C., Palanque, P., Barboni, E., Deleris, Y., & Gris, C. (2018). Engineering issues related to the development of a recommender system in a critical context: Application to interactive cockpits. International Journal of Human-Computer Studies. https://doi.org/10.1016/j.ijhcs.2018.05.001.

[6] Adebayo A, Agbola I, Ayangbade A, & Obajimi O. (2015). Bank Products Recommender. The International Journal Of Engineering And Science (IJES), 4(7), 2319–1805. Retrieved from www.theijes.com

[7] Emil, E. (2019). Why graph database could be key to addressing financial services challenges. Financial IT. Retrieved from https://financialit.net/blog/financial-services/why-graph-database-could-be-key-addressing-financial-services-challenges.

[8] Felfernig, A., Jeran, M., Stettinger, M., Absenger, T., Gruber, T., & Haas, S. (2015). Human computation based acquisition of financial service advisory practices. Proceedings of the 1st International Workshop on Personalization & Recommender Systems in Financial Services

[9] Felfernig, A., & Stettinger, M. (2015). Conflict Management in Interactive Financial Service Selection. Organizational Support, 3.

[10] Musto, C., Lops, P., Gemmis, M., Lekkas, G., & Semeraro, G. (2015). Personalized finance advisory through case-based recommender systems and diversification strategies. Decision Support Systems, 77, 100-111

[11] Abbas, A., Bilal, K., Zhang, L., & Khan, S. (2015). A cloud based health insurance plan recommendation system: A user centred approach. Future Generation Computer Systems Journal, 43(44), 99–109

[12] Guo, G., Zhang, J., & Yorke-Smith, N. (2015). Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems. Knowledge-based Systems 74, 14–27.

[13] Gigli, A., Lillo, F., & Regoli, D. (2017). Recommender Systems for Banking and Financial Services. In Proceedings of RecSys Posters, Como, Italy.

[14] Reilly O. (2019). Why Graph databases? Retrieved March 14, 2019, from https://neo4j.com/why-graph-databases/