Journal of Technological Information, Management & Engineering Sciences (2020),1(2), 36-52

Qureshi and Zahra

Optimizing a Recommender System for B2B Digital Marketplaces Based on Collaborative Filtering and Graph Networks

[1] **Dr. Saad Qureshi**

[2] **Fatima Zahr**

How to cite this article:

Qureshi, S., & Zahra, F. (2020). Optimizing a Recommender System for B2B Digital Marketplaces Based on Collaborative Filtering and Graph Networks. *Journal of Technological Information, Management & Engineering Sciences, 1(2), 36-52.*

Received	Accepted	Published
22 July 2020	07 September 2020	27 November 2020

Abstract

Business-to-business (B2B) Digital Marketplaces are those that link businesses with suppliers, products, services, and business opportunities. Unlike consumer platforms, B2B platforms typically consist of few interactions, high dollar transactions, long sales cycles, complicated product attributes, organizational purchasing positions, negotiated pricing, and buyer-supplier-cataloged item relationships. These features can make the task of optimizing recommendation quality challenging with a single standard approach. In this paper, a hybrid recommender-system model fusing collaborative filtering with graph-based representation learning is proposed to enhance the relevance, discovery, and operational value in B2B marketplaces. Collaborative filtering looks at the patterns of buying–selling relations between buyers and items and graph networks are used to express multi-relational relations between organizations, items, suppliers, categories and transactions. The architecture proposed is data preparation, implicit-feedback modelling, graph construction, candidate generation, ranking, business-rule constraints and online monitoring. It covers metadata and graph connectivity for cold-start, and it includes temporal validation, ranking metrics, diversity, supplier exposure and commercial outcomes in evaluation. Since the paper does not include any market data, it does not contain fabricated experimental results, instead it proposes a research design and proposes some hypotheses. The anticipated delivery is a realistic framework of the benefits that graph-enhanced collaborative filtering can provide over basic approaches, in terms of measurable improvements in the metrics of explainability, privacy and marketplace fairness. The paper argues that recommender optimization must be a "whole-of-system decision problem," considering relevance, discovery, commerciality, system latency, and treating marketplace participants as appropriately.

Keywords: *B2B marketplace, recommender systems, collaborative filtering, graph neural networks, implicit feedback, link prediction, ranking, cold start, supplier discovery.*

1. Introduction

Digital marketplaces have become key infrastructure of business procurement and wholesale distribution, industrial sourcing, and business services. They cut

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

down on search costs by consolidating suppliers and catalog information, but the magnitude can complicate the process of discovering products. A buyer may find themselves dealing with thousands of products whose specifications overlap, which come in many different qualities, in many different available quantities, for negotiated pricing, and have a wide variety of fulfillment conditions offered by the suppliers. This can be alleviated by recommender systems which focus on products and suppliers that fit a buyer's operational needs.

There are some basic differences between B2B recommendation and consumer recommendation. The purchasing events occur infrequently but can be costlier, orders may be placed a few months ahead of time, there may be technical interchangeability but not substitutability, and purchasing decisions may be made by teams, not individuals. When buying in the past, a buyer's past actions may have been due to contracts, inventory policies, approved vendors or temporary inventories, and not a personal preference. As a result, a recommendation model should employ contextual and relational information as opposed to simply viewing each past purchase as a vote of appreciation.

A popular and widely-used technique is collaborative filtering, which learns from the patterns of interaction between users and items. It can reveal hidden preferences even if the product descriptions are not unambiguous. But the interactions are thin and new purchasers or products diminish its usefulness. Graph-based methods offer a complementary approach and visualizing marketplace entities and their relationships as a graph. There are examples of graphs that link organizations with categories, products with suppliers, products with compatible components, and buyers with previous purchases. Multi-hop relationships can be used to infer relevance that is not apparent in a graph representation of a buyer–item matrix.

In this paper, a hybrid architecture is proposed, which integrates collaborative filtering and graph networks. It does not consider recommendation as a one model, but as a pipeline: candidate generating, ranking, business constraints and elimination, monitoring model quality and marketplace outcomes. The goal isn't

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2020),1(2), 36-52

to make unverified claims, but to outline a repeatable framework to be tested empirically.

1.1 Research Objectives

This study aims to fulfill four goals: (1) to identify the requirements that make B2B recommender systems different from consumer recommender systems; (2) to design a hybrid collaborative-filtering and graph-network architecture; (3) to define the methodology for measuring the quality of the ranking and its commercial utility; and (4) to consider cold-start behaviour, diversity, supplier's exposure, privacy, and deployment constraints.

1.2 Research Questions

The question posed in the paper is: What strategies should be used to combine collaborative filtering and graph networks for B2B recommendations? Can the structure of the graph be used to alleviate interaction sparsity and the cold-start problem? What are the right metrics to assess the quality of recommendations in business marketplaces off-line and on-line? How does optimization enhance the relevance of the buyers if it does not end up focusing the visibility on a few suppliers?

2. Literature Review

The most frequently mentioned types of recommender systems are collaborative filtering, content-based, and hybrid. Collaborative filtering is based on patterns of users and items to predict preferences. Memory-based methods calculate similarity between buyers or items, while model-based methods learn the representations of the objects or directly predict the probability of interaction. One special approach gained significant popularity due to the mapping of users and items to a shared latent space, and the possibility of predicting the preferences of unobserved users or items from sparse interaction data: matrix factorization (Koren, Bell, & Volinsky, 2009). In marketplace context, implicit feedback such as clicks, search, product views, quote requests, cart additions, purchases and repeat orders are more available than explicit ratings.

Implicit feedback is not clear. A product view might indicate interest, but might also be a result of incidental viewing, price comparison or price checking. A

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

purchase can be a good sign of commercial relevance, but it could be due to a contract or only because the product was available. Hu, Koren, and Volinsky (2008) proposed a confidence weighted implicit feedback recommendation method where they assign weights to observed interactions based on the confidence associated with the interactions. For a B2B system, interaction types can be given business significance and adjusted for position bias, repeated interactions, and the time since the interaction.

Neural collaborative filtering is an extension of latent-factor methods that models the interaction between item and user embeddings with a neural network (He et al., 2017). These can be used to encode more complicated patterns, but with proper regularization and evaluation. Sequential recommenders can also account for the sequence of events, such as when users move from browsing to requesting a quote to making a purchase or come back for more items. B2B however, can have longer buying cycles, and so sequences length and context need to be based on the buying cycle, not just consumer browsing.

A graph-based recommendation is a representation of entities and relations as nodes and the edges. Graphs can visualise the interactions between buyers and products as well as the interactions between suppliers and products, product category, compatible components, geographic availability, and organisational affiliations. Graph convolutional methods combine the information of adjacent nodes to form local structure representations. To simplify graph convolution in CF, for instance, propagation and combination of embeddings in interaction-graph neighborhoods are implemented in LightGCN (He et al., 2020). Knowledge-graph methods can use typed relations and descriptive attributes, enabling the model to make inferences about several types of relation.

Learning to graph does not make it a good learning. Dense or noisy connections can convey irrelevant information, while large graphs can yield storage, computation and freshness problems. Historical concentration can be encoded in graph structure – if a supplier is popular, he gets more interactions, he is made more visible, he gets even more interactions. Exposure-aware evaluation could have the opposite effect of a graph model, in that it may serve to amplify

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

popularity bias. Furthermore, graph-based explanations need to be carefully examined—linking a buyer and a product may be a way to convey relevance without implying that the recommendation is causally relevant.

To do offline evaluation for top-N recommendation one can rely on metrics like precision at K, recall at K, normalized discounted cumulative gain (NDCG at K) and mean reciprocal rank. However, what is observed in historical logs are only the exposed and selected entities based on an older ranking policy. A model could look weak because they were not all presented with products that were relevant to them, and strong because they were over-exposed to popular products. As a result, temporal splits, unbiased logging (where possible), online controlled experiments, and business metrics are required to complement the traditional offline tests. The selection of the type of split, negative sampling, and candidate set can significantly influence results (Rendle et al., 2009; Steck, 2010).

3. Requirements for B2B Marketplace Recommendations

The student will describe and identify the preferences related to organizational and role-based preferences.

A B2B marketplace should differentiate the organization from the individual user and the purchasing role. The same person can be involved in a procurement process for various things at the same time, as a procurement officer, engineer, finance approver and warehouse manager. The recommendation for a technical user may focus on compatibility and specifications, and for a procurement user, it may focus on approved supplier, price, lead time, and contract terms. Approved categories can be summarized at the account level, historical purchasing trends and procurement constraints can be summarized at the account level, and the model can represent users nested within organizations. Access control should not allow private purchase history to be leaked to another organisation.

3.2 High-Value, Low-Frequency Events

Many B2B transactions involve more than one buyer or seller and are high value. If there is low count of interactions, don't assume low interest; buyers might buy specialized equipment once every few years. The system should be able to make recommendations when there is a limited amount of interaction history between

[1] Department of Computer Science, Air University, Islamabad, Pakistan
[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

products and suppliers, based on product attributes, supplier reliability, compatibility relations, and the behavior of the products at category level. Recency should relate to product lifecycle and replenishment cycle; a product purchased a year ago can still be relevant in the case of industrial equipment but not perishable stock.

The commercial and operational constraints are discussed.Commercial and operational constraints are discussed.

The quality of recommendations is limited by the business feasibility. Products might not be available, limited to certain regions, and not be compatible with current systems, or not meet an organization's approved-vendor standards. Hard constraints should be applied after candidate generation and not after final display in the serving layer. Ranking policy should include soft objectives (for example, margin, diversification of suppliers, and strategic category discovery) as a clear, open part of the ranking process, not as a component of relevance. Safety, legal, contract, or buyer-specific eligibility requirements cannot be compromised for commercial reasons.

3.4 Multiple Objectives

Optimizing the marketplace can mean optimizing for click-through, quote requests, conversion, gross merchandise value, repeat purchase, buyer satisfaction and/or supplier participation. These outcomes are not all equal. Optimising clicks alone could be a way of pushing good pictures or selling products at low prices that do not meet the procurement requirements. A multi-objective system should include a primary metric and guardrail metrics like complaint rate, qualified conversion, cancellation rate, return rate, supplier exposure, latency. Business stakeholders should be involved in determining the relative weights and reviewed periodically.

4. Proposed Hybrid Architecture

The proposed system includes six stages: data ingestion, construction of features; collaborative-filtering-candidate generation; graph-based representation learning; hybrid scoring and ranking; policy filtering and presentation; and feedback monitoring. The architecture provides a separation of the graph model and

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

collaborative model so that they can be compared independently, but merged only to gain complementary value.

The basics of Data Ingestion and Interaction Modeling.

The system captures the allowed events like product impressions, clicks, searches, saves, quote requests, cart additions, orders, repeat purchases, and explicit feedback. For each event, a time stamp, organisation identifiers, user identifiers, item or supplier identifiers, context and (where possible) exposure position should be included. Deduplication and bot and test traffic and abnormal events are detected. The data model not only captures positive interactions, but also all the items that were technically eligible and available to a buyer and are exposed to them, which helps prevent false negatives for every item that was not clicked.

Each event type can be given a confidence weight for interaction. In some instances, an order received may be more confidence inspiring than an impression of a product, such as in categories where procurement is negotiated, an order request can be a strong signal. It is important to note that the weights should be adjustable according to empirical considerations instead of being assumed as universal. One way to model time decay is to use the function $w(t) = \exp(-\lambda \Delta t)$, where Δt represents time since the previous observation and λ is determined by the category's expected lifetime. A category-specific decay is more desirable than one global decay if there are vastly different replenishment cycles within the network.

4.2 Collaborative Filtering Component

Collaborative-Filtering learns latent vectors for organizations, user-role and products. The matrix factorization representation sees the score for a particular buyer u and item i as the dot product of the embedding p_u and q_i, possibly with the addition of bias terms and context features. A confidence-weighted objective can take the observed events as evidence of positive preferences and use the sampled unobserved events as evidence of lower-confidence preferences as implicit feedback. Samples should be taken to honor exposure and eligibility: if an item is not available to the buyer, it should not be considered a meaningful negative.

[1] Department of Computer Science, Air University, Islamabad, Pakistan
[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

The collaborative model is passed to candidate generation, which returns a set of top-scoring items. It may be enhanced with popular-in-category products, recently added products, and products that have similar attributes to those already purchased. The size of the candidate set should be sufficiently large to ensure a diverse ranking, but not so large that serving it is too slow. This is the case where collaborative filtering will work best, i.e. for well-known products for which there is meaningful interaction history, and the buyer is not new.

4.3 Graph Construction

The graph is heterogeneous as it has more than one node and edge type. The nodes can be organizations, user roles, products, suppliers, categories, brands, compatible components, locations, and transactions. Edges can be purchased, viewed, quoted, supplied-by, belongs-to-category, compatible-with, substitutes-for, or available-in-region. Every relation should have a well-documented meaning and date of creation, if applicable.

The graph should not just tie everything together. Quality checks, provenance, restriction and time-validity rules are needed for edges. A terminated supplier relationship should not continue to exist for an unlimited period, for instance; compatible edges should be provided by dependable technical sources, for example; or private commercial terms should be provided only to authorized components. The graph is built using data at the recommendation time stamp to avoid future-event leakage when training and testing.

4.4 Graph Network and Representation Learning

To update a node representation, each node in a graph network combines information from the neighboring nodes. A generic layer is defined as $h_r^{(l+1)} = \sigma(W_0 h_r^{(l)} + AGG(\{W_r h_u^{(l)} : u \in N_r(v)\}))$ where h_r is the representation of node v, $N_r(v)$ is the neighborhood of v under relation r, W are learned transformations, and σ is a nonlinear activation. The model can differentiate a purchase edge from a compatibility edge by a supplier edge, using relation-specific transformations.

In collaborative recommendation, one can build a simplified graph convolution model where the buyer and the item are fed through the buyer–item interaction graph and then can be fused together across various propagation depths. A

[1] Department of Computer Science, Air University, Islamabad, Pakistan
[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

relational graph neural network or knowledge-graph embedding model can be used to embed typed relationships in the case of a heterogeneous marketplace. This should be based on graph size, edge quality, serving latency and the semantics of the relations. The graph component can be used to score items for the buyer similarity-based approach, or to pass features to the downstream ranking model.

4.5 Hybrid Scoring and Ranking

The proposed hybrid score is a combination of collaborative filtering, graph relevance, content similarity and allowed contextual signals. Its simple combination is $s(u,i) = \alpha s_CF(u,i) + \beta s_G(u,i) + \gamma s_C(u,i) + \delta s_X(u,i)$, with each term corresponding to a collaborative, graph, content, and contextual score. The weights can be trained using a validation set or they can be fed into a ranking model. All component scores should be scaled or calibrated so that the sum of the components is not warped by the different scales.

A more flexible approach is to use an impression-level learning-to-rank model. They can include collaborative score, graph score, category match, specification compatibility, supplier reliability, estimated lead time, buyer-account preference, and recency. The target for the ranking may be a weighted result based on factors such as qualified quote requests and purchases that are greater than impressions. Training labels and business weights should be noted and assessed in relation to downstream outcomes rather than being over-optimised in one event.

Business Rules, Diversity and Re-Ranking 4.6

The system eliminates items that do not meet the criteria, and can re-rank using diversity-aware re-ranking. Diversity can help to eliminate nearly identical responses and provide the buyer with alternatives from other relevant categories or suppliers. One supplier may not be able to hold all the top roles if there are other suppliers with similar qualifications. But diversity shouldn't be a deciding factor in determining which products go on the short list, and which do not. The final list should maintain some minimum level of relevance and eligibility while at the same time being discoverable and workable.

[1] Department of Computer Science, Air University, Islamabad, Pakistan
[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2020),1(2), 36-52

5. Research Methodology

This paper adopts a design-oriented research approach of stating a system architecture, identifying measurable hypotheses and presenting a protocol for comparative evaluation. This is not an actual empirical experiment, as no original marketplace data was provided. No intentional skirting of the truth with numerical performance claims. Use of a marketplace dataset, which must be a licensed dataset that has been reviewed for privacy, or a documented benchmark with limitations clearly documented.

The data set and experimental unit are the same.The data set is identical to the experimental unit.

The dataset preferred is event logs, item metadata, supplier information, catalog taxonomy, organization-level context and timestamps. The experimental unit is a buyer–item opportunity at a particular recommendation time. It is important to document what was eligible and exposed, since an item that was not exposed cannot be interpreted as an actual rejection. If a log of exposure has not been kept, the study should mention the limitations of implicit negative sampling.

5.2 Temporal Data Splitting

Training, validation and testing sets should be kept apart in time. The models are trained with earlier interactions, hyperparameters and blend weights are tuned in a second window and the final system is evaluated in the last window that matures. All graph edges, item features and account summaries shall be recreated as they would have been presented for each evaluation time. Random splitting can also be likely to leak future behavior into training and exaggerate performance, particularly where repeat purchases and consistent supplier relations result in a high level of temporal dependence.

5.3 Baseline Systems

A minimum of, compare the hybrid with the following: (1) popularity and category-popularity ranking; (2) item-item collaborative filtering; (3) matrix factorization or implicit-feedback collaborative filtering; (4) a content-based recommender; (5) a graph-only recommender; and (6) the proposed collaborative-filtering-plus-graph model. Wherever possible the same candidate

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

eligibility rules should be applied, and the same test period. The component that is eliminated should be one of the components in a graph, collaborative component, content features, or diversity re-ranking. Ablation experiments should isolate one of these components at a time and examine its contribution.

5.4 Evaluation Metrics

Precision@K is the ratio of the number of relevant items retrieved out of the top-K items returned, and Recall is the ratio of the number of relevant items retrieved out of the total number of relevant items. NDCG@K will be rewarded by relevant items having higher positions in the list and can even have a graded relevance. Mean reciprocal rank gives more weight to the rank of the first relevant item. The coverage captures the fraction of the catalog/catalogue or supplier base that is exposed to recommendations, and intra-list diversity captures the difference between the recommended items. Novelty can be measured as recommendations appear less obvious items, but novelty does not necessarily mean commercial use. Qualified quote-request rate, conversion rate, order value, repeat purchase, cancellation, returns, and time to successful product discovery are important business metrics. The metrics should be easily attributable, and broken down into categories, buyer tenure, order value, and new vs established suppliers when sample sizes allow. Also significant are system metrics like p95 latency, graph refresh delay, and serving availability. Ranking improvements offline cannot be taken as a given that it will generate revenue improvements without evidence online.

5.5-Online Testing and Statistical Validity

If feasible from an operations perspective, an A/B test can be used to randomly select an eligible organization or session in order to create a baseline and treatment system. Randomization should be done at a level to minimize contamination, such as organization-level assignment when more than one user are involved in purchasing assignments. The main metrics, guardrails, duration of test and minimum detectable effect, stopping rules should be established prior to the test. The experiment should follow a relevant buying cycle for the duration of the test – shorter tests may not capture delayed B2B conversions. When

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2020),1(2), 36-52

randomization is not feasible, a quasi-experimental method with careful justification may be employed, with limitations documented.

6. Proposed Algorithmic Workflow

Step 1: Establish the recommendation objective, eligible users, scope of the catalog, labels of the desired outcome, and business constraints.

Step 2: Consume and validate buyer, product, supplier, interaction and catalog data that is timestamped.

Step 3: Construct time-aware interaction matrices and time-weighted implicit feedback examples.

In Step 4, train a collaborative-filtering model and produce a list of candidate products for each purchaser.

Step 5: Build a heterogeneous graph with only relationships at the recommended time stamp.

Training a graph network or graph collaborative-filtering model, and computing relevance scores for buyers and items.

Step 7: Combine the candidates and normalize the collaborative, graph, content and contextual features.

Step 8: Train/tune a hybrid ranking model from validation data and leakage-safe examples.

Step 9: Enforce strict criteria, availability filters, diversity re-ranking and supplier-exposure guardrails.

Step 10: Analyze on an undisturbed temporal test set based on ranking, coverage, diversity, business and latency.

Step 11: Perform the ablation, cold-start, segment, and robustness analysis.

Step 12: Implement from a trial basis, logging, monitoring, roll back, and retraining periodically for review.

The expected outcomes and analysis plan. The expected outcomes and analysis plan.

The primary hypothesis is that the hybrid system will be able to increase top-K relevance for existing buyers using the interaction patterns as well as graph relationships, and increase cold-start recommendations when new products or

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2020),1(2), 36-52

suppliers are available and have relevant metadata and graph relationships. These are tentative conclusions and not facts. If the graph edges are noisy, the marketplace does not have much graph data or the collaborative and graph models learn comparable data, then gains can be constrained.

All baselines should be compared in a results table for Precision@K, Recall@K, NDCG@K, catalog coverage, supplier coverage, diversity, qualified conversion, and latency. Results should provide confidence intervals if possible, and results should be segmented by buyer tenure, by category and by product novelty. The report should include the number of candidates generated, the negative sampling method used, the definition of relevance, and time limits for the relevance definition because these design decisions have an impact on how the metrics can be interpreted.

Yes, in an ideal family; Yes, in a good family; Yes, in a family with some issues; or No, in a problematic family.

Popularity / category popularity Strong simple baseline; robust for sparse histories Limited personalization & novelty

Weakly supervised Learns behavior similarity and repeat patterns Cold start and sparse interactions

Graph-only model: Multi-relational connections and neighborhood context: Sensitive to graph quality and computation

Hybrid CF + graph Can combine behavioral & relational signals More complex; need for ablation & governance

The analysis of cold start and sparse data.7.1 Cold-start and Sparse-Data Analysis.

Cold-start evaluation should be split among new buyers, new products and new suppliers. The graph model could be used for new products, with relationships between suppliers, categories, specifications, compatibility and region. In the case of new buyers, it may be possible to make recommendations based on organization-level category needs and explicit procurement context, if these signals can be obtained and approved. Content-based and category-popularity baselines should be used for results comparison. A graph model cannot make

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

meaningful inferences of preferences based on weak or nonexistent relationships, and therefore it requires fallbacks.

To understand the concept of Diversity and Supplier Exposure.

The successful recommender must duplicate the most popular suppliers. The proportion of suppliers covered, the concentration of the exposure, and recommendation share by supplier size can provide an indication of the systematic preference for incumbents in the model. The study should assess relevance and exposure together as both are not only important, but also because maximizing both without taking into consideration the needs of buyers can decrease the utility of the study. The use of any exposure intervention should be recorded as a policy decision and assessed for adverse reactions.

8. Discussion

A hybrid architecture that combines the advantages of collaborative filtering and graph learning is proposed. The concept of Collaborative Filtering aims to detect repeated behavior, and can be useful for finding candidates from a big catalog. Graph networks are used to represent relationships that cannot be captured in the context of direct relationships between buyers and items, such as product compatibility, the structuring of categories, supplier relationships, or the organization in which products exist. The two can be combined to offer a more complete picture of buyer intent, especially if interactions are limited, but metadata remains solid.

The integration does not come without a price tag. Reliable entity resolution, relation governance and infrastructure are necessary to construct and refresh the graph. The model might be less easy to explain than item-item similarity, or just a popularity baseline. Errors can also be spread across a graph; for example, if a compatibility edge is mislabelled or a compatibility supplier relationship is outdated, the recommendations will be impacted in more than one record. Production systems should, therefore, have provenance and confidence in graph edges, enable the deletion and correction of graph elements, and quantify the effect of graph changes.

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

An optimization process must have multiple objectives for B2B marketplaces. Buyer value is related to relevance, compatibility, reliability, lead time and ease of procurement. Other factors, such as conversion, retention, supplier inclusion and effective inventory identification, are also important factors affecting marketplace value. These objectives may intersect. For instance, a ranking designed to get the highest amount of orders in the short term could hinder the discovery of new suppliers or category growth. A defensible system puts these trade-offs on the table explicitly, with strict standards for eligibility and compliance, and softer business measures with agreed guardrails.

8.1 Privacy, Security, and Responsible Use

The histories of interaction within the market can uncover sensitive commercial information, including sourcing strategy, supplier relationships, and planned purchases. Data access should be done in a least-privilege way and model training should not involve the exposure of the behavior of one organization to another. Identifiers should be anonymized if possible, retention should be to known purposes, and access to graph neighborhoods should be restricted by organization permissions. Membership inference, data leakage and unwanted inference from explanations of recommendations are some aspects to be considered in the security review.

There must be transparency with respect to the role of the system when it comes to responsible recommendation. Recommendations must not suggest certification, compatibility or a guaranteed delivery date unless backed by authoritative data. Human procurement professionals should be able to review specifications and change recommendations. Supplier ranking should be measurable and, in particular, auditable if paid placement or commercial incentives are key factors in determining the ranking.

The limitations and future research of the proposed study are discussed in Section 8.2

The biggest drawback is that the paper outlines an architecture and evaluation plan, but has not implemented it on real marketplace data. The real value of the graph learning will rely on graph density, catalog quality and user behavior,

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

category structure, and operational factors. Observed purchases may not accurately reflect buyer preference, and historical interaction data are influenced by previous interactions and commercial policies. Offline metrics alone are not enough to determine causal effects on revenue or satisfaction.

The framework needs to be studied in other B2B sectors such as industrial components, wholesale goods, office supplies and business services. Should be able to compare different heterogeneous graph networks, LightGCN-style models, graph embeddings, and simpler relational features. Future research may explore graph updates over time, user-understandable recommendation steps, privacy enabled collaboration, counterfactual algorithms, and supplier-exposure restrictions. For categories with long procurement cycles, long-running experiments might be required to measure retention and repeat purchasing.

9. Conclusion

This paper suggested an hybrid recommender-system framework for B2B digital marketplaces, namely Collaborative Filtering and Graph networks. This design features interaction based candidate generation, heterogeneous graph representation learning, hybrid ranking, eligibility filters, diversity-aware re-ranking, and systematic monitoring. It is designed for B2B transactions with their unique traits like infrequent engagement, large ticket sales, organizational needs, product compatibility, supplier dynamics, and business restrictions.

The main point of the central conclusion is methodological: the quality of recommendations should be measured as a whole system and not using one metric offline. Only when significant gains over simpler baselines have been demonstrated through temporal testing and controlled evaluation should a hybrid approach be used. In B2B marketplaces, the proposed framework offers a structured basis for evaluating these, and for providing more meaningful, understandable and practically viable recommendations.

References

Burke, R. (2002). Survey and experiments on hybrid recommender system. User Modeling and User-Adapted Interaction, 12, 331–370. https://doi.org/10.1023/A:1021240730564

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). Neural collaborative filtering. In the 26th International Conference on World Wide Web (WWW), 1997. 173–182). https://doi.org/10.1145/3038912.3052569

He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. (2020). LightGCN: Recommendation with Simplified and Powered Graph Convolution Network. We present the results of our work in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 639–648). https://doi.org/10.1145/3397271.3401063

Hu, Y., Koren, Y., & Volinsky, C. (2008). Implicit Feedback based Collaborative Filtering. The eighth IEEE International Conference on Data Mining (IEEE ICDM 2008) took place in 2008. 263–272). https://doi.org/10.1109/ICDM.2008.22

Koren, Y., Bell, R., & Volinsky, C. (2009). Technique for recommender systems: matrix factorization. Computer, 42(8), 30–37. https://doi.org/10.1109/MC.2009.263

S. Rendle, C. Freudenthaler, Z. Gantner, & L. Schmidt-Thieme. (2009). BPR: Bayesian personalized ranking with implicit feedback. The Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI 2025) did not feature any papers on these topics. 452–461).

Steck, H. (2010). Training and testing of recommender systems on data missing not at random. In the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '91), 1991. 713–722). https://doi.org/10.1145/1835804.1835895

Wang, X., He, X., Wang, M., Feng, F., & Chua, T.-S. (2019). Neural graph collaborative filtering. In SIGIR '10: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2010. 165–174). https://doi.org/10.1145/3331184.3331267

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). A survey and new perspectives on deep learning based recommender systems. ACM Computing Surveys, 52(1), Article 5. https://doi.org/10.1145/3285029

[1] Department of Computer Science, Air University, Islamabad, Pakistan

[2] Department of Software Engineering, International Islamic University, Islamabad, Pakistan

Copyright (c) 2020 Author(s)

This work is licensed under a Creative Commons Attribution 4.0 International License.