Complementary Recommendations: A Brief Survey

Hang Yu[1]*, Lester Litchfield[1], Thomas Kernreiter[1], Seamus Jolly[1], Kathryn Hempstalk[1]
[1]*Trade Me Ltd, Wellington, New Zealand*
Orcid: 0000-0002-3087-744X, 0000-0002-1687-0508, 0000-0002-3716-8603, 0000-0001-7811-3038, 0000-0002-4023-9812

***Abstract*—Driven by the rapid development of global e-commerce, recommender systems have become a hot topic across both industry and academia because they offer a potential source of business revenue. A wide range of recommendation solutions have been developed and they can be classified as substitute and complementary: substitute recommenders offer similar items to the source item, complementary recommenders suggest items that are dissimilar to the source but are often sold as a companion item or service (such as a mobile phone (source) and case (complementary)). Both types of recommendations demonstrate unique values in specific application domains. However, the research and development of complementary recommendations still remain sparse when compared with substitute recommenders. This paper presents a brief survey on existing solutions for complementary recommendations. Our work summarizes the existing research activities and explores open questions in this field by discussing three aspects including the identification of the ground truth for complements; the recommendation models; and evaluation datasets and metrics. To the best of our knowledge, this work is one of the few surveys that provide particular insights on complementary recommendations in recent years.**

Keywords*—*E-commerce, recommender systems, complementary recommendations

I. Introduction

Recommender systems are increasingly attracting the attention of both industry and academia because of the potential value generated in e-commerce applications [1]. For recommender systems, the ultimate target is to stimulate the purchasing activities of potential customers by retrieving the items that catch their personalized interests among the overloaded information. With this goal, such systems can be classified as substitute and complementary recommenders: substitute recommenders offer similar items to the source, complementary recommenders suggest items that are dissimilar to the source but are often sold with it as a companion item or service (such as a mobile phone (source) and case (complementary)).

Compared with the deep and mature research outcomes on substitute recommenders, there still exist research gaps and opportunities for complementary recommenders. However, the research and development of complementary recommenders face unique challenges that require specific methodologies to address; the most typical challenge resides in complements discovery that is non-trivial for the system. Unlike substitutes that could be explicitly identified as similar products interchangeably viewed or purchased, a large number of complementary products are loosely bound by latent principles that are difficult to capture effortlessly. As a result, the relationships might be inappropriately defined as opposed to the ground truth, which imposes extra pressure on the model training and tuning. Correspondingly, from the perspective of modeling, the high complexity of complementary relationships may limit the direct application of some substitute recommendation models such as the ones that rely on simple similarity measures considering the possibly long distance between two complements in feature space [2].

This work supported by Trade Me Ltd, New Zealand.

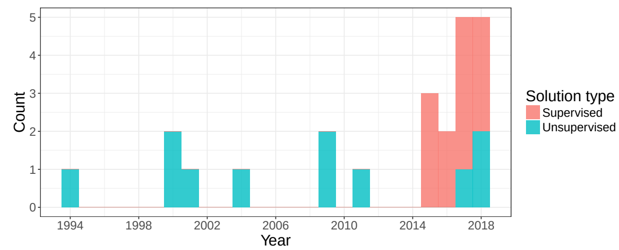

Fig. 1. Publications over year

The above challenges need to be addressed considering the high value of complementary recommendations [3]. This paper surveys 23 publications on complementary recommendations published between 1994 and 2018 including two publication sources that are conference proceedings and journals. The time distribution of all the publications, displayed in Fig. 1 and surveyed in this paper, shows that the research on complementary recommendations remains sparse before 2014 but demonstrates an increased popularity afterwards. Conference publications account for 87% of all papers surveyed.

In comparison with other surveys that mainly focus on substitute recommendations [4]–[9], we aim to contribute specific insights on complementary recommendations to future researchers in the following aspects:

1) The identification of the ground truth for complementary products.
2) Currently available complementary recommendation models.
3) Currently available evaluation data sets and metrics for complementary recommenders.

The rest of the paper is organized as follows: in Section II, the identification of the ground truth for complementary products is introduced. Next, review of modeling including unsupervised and supervised learning is presented in Section III, followed by the summary of the typical evaluation data and metrics in Section IV. Finally, the conclusions are presented in Section V.

II. Identifying the Ground Truth for Complementary Products

Identifying the ground truth for complementary products plays a significant role in modeling complementary relationships. Compared with substitute products that can explicitly be defined by interchangeability or similarity, identifying the complements for items requires more effort because complements are usually loosely associated by various implicit principles.

In economics, a complement is defined as an item that is usually consumed with another. Correspondingly, the cross elasticity of demand is negative, which means that price drop for one item increases demand for its complements [3], [10]. In other words, the purchase of the complements of one item is positively related to that of the item itself. Guided by this theory, identifying complements based on frequent co-purchase becomes the first rule of thumb for the vast majority of the research papers reviewed. Within this context, data sources that potentially carry complementary items could be further classified by the cross-interaction between event time and object granularity as summarized below:

1) Simultaneously purchased items [2], [11]–[16].
2) Simultaneously purchased categories [17], [18].
3) Consecutively purchased items [12]–[15], [17], [19].
4) Consecutively purchased categories [10], [17], [18].

However, co-purchase is not sufficient to infer a complementary relationship as it can be further characterised by considering the following principles:

1) Asymmetry [2]. The co-purchase of a source item and its complements is asymmetric, which means the purchase of the complement usually occurs as a result of that of the source item rather than the other way around. One typical example is that people usually purchase power banks for their mobile phones however they rarely buy mobile phones for their power banks. This needs to be considered when graph-based methods are applied.
2) Transient [20]. Due to the transition of transactional behaviours, the association between the complements and source item decays over time. For example, the products simultaneously purchased with a mobile phone are more likely to contain the corresponding complements such as power banks, cables, and screen protectors, compared with the ones purchased one month later when user interest has already shifted. Therefore, the item consecutively purchased after the source item can be identified as a complement only when the time interval complies with certain scenario-specific thresholds. The threshold needs to be carefully selected when sampling complementary pairs from sessions.

The above-mentioned principles help mining the characteristics of complementary relationships so that the economic definition can be interpreted into practical operations. This step builds the foundation for modeling complementary relationships as discussed in the next section.

III. Modeling Complementary Relationships

Given the ground truth for complements, complex complementary relationships can be modeled in various ways. In this section, existing models adopted for complementary recommendations are summarized. Solutions are clustered into two groups that are unsupervised and supervised learning so as to show the development and applications of different methods.

A. Unsupervised Complementary Recommender

As presented in Fig. 1, complementary recommenders based on unsupervised learning were preferred at the beginning. Among all, association rules [21] are firstly applied in market basket analysis to explore the relationships among products including complementary sets. In the basic implementation of association rules, one product is regarded as the complement of another if the their co-occurrence is greater than two thresholds called minimum support and minimum confidence. Regardless of the simplicity and interpretability, association rules are vulnerable to parameter selection which often leads to a large number of irrelevant connections. Thus, many variants of association rules have been proposed to increase the relevance of products. Such algorithms include rule pruning [22], [23], interestingness measurement [24], association rule networks [25], and the K-RecSys [26] combining collaborative filtering and association rules.

Regardless of the interpretability, association rules are not appropriate for large transaction datasets that contain complex relationships. One solution for this issue is fast search such as Frequent Pattern (FP) growth [27] based on FP trees that are a compact structure constructing the global frequent patterns of items. Based on the FP tree, the FP growth is able to rapidly searches for all frequent patterns with linear complexity. Alternatively, product networks and community detection are combined [11]. Firstly, network edges are pruned by setting a threshold defined by the co-purchase frequency between items so as to eliminate the noise caused by coincidental purchases. Next, communities are detected in an unsupervised manner that optimizes the modularity.

The solutions mentioned above explore complementary relationships based on co-purchase frequency. In some cases, complementary relationships cannot be accurately detected purely based on co-purchase due to the noise introduced by the transactions occasionally mixed by both complements and substitutes or even pure substitutes. To increase the precision of identification, finding complementary products by jointly considering the effect of complements and substitutes has become an alternative direction. The SC tagging [10] algorithm is one of the early attempts following this direction. The first step of SC tagging is to classify user activities as the comparing stage and reviewing stage that potentially contain substitutes and complements, using user events including purchase, add to cart, and fail in bidding. Next, the complementary relationship is determined by a metric that compares the complementary level against the substitute level. A similar statistical metric that measures complementary relationship is the Complementary Score [28]. Compared to the symmetric SC tagging, one major improvement of this method is that it builds complementary and substitute relationships as a directed graph by taking into consideration the sequential information of items, which increases the fidelity.

With the recent development of feature representation enabled by deep learning techniques, the capacity of unsupervised statistical models is envisaged to be largely improved. One of the typical model is the Baskets and Browsing to Vector (BB2Vec) [12]. Being an extension of the prod2Vec model [29], BB2Vec also utilizes the idea of Word2Vec to generate item embeddings. However, compared to its predecessors, this model joins basket data and browsing history as a multi-task learning problem that is followed by a learn-to-rank training process. In this way, the shortcoming of insufficient co-purchase data is alleviated. As a result, complementary products yield a high score for the dot product of their embeddings.

Quite generally, most of the unsupervised methods model complementary relationships by discovering similarities. As a result, their applications are limited because of the high complexity of complementary relationships that demand advanced modeling techniques.

B. Supervised Complementary Recommender

Although supervised approaches may suffer from the noise introduced by human-labeled ground truth, they are capable of modeling relationships with high complexity using the advanced techniques such as deep learning. Therefore, supervised complementary recommenders have become the preferred solutions in recent years. Considering the large number of supervised models, we cluster them into subsections based on their mutual similarities.

1) Visual model: Motivated by the success of deep learning in image classification and it's unique ability to represent users and items as embeddings, several visual recommender systems are proposed as pioneers to address the complementary recommendation task for e-commerce. The *visual and relational recommender system* [17] is an early attempt to model relationships among items based on their image features. The authors use large-scale co-purchase data from *Amazon.com* as the ground truth for complementary products. Using visual features extracted from pre-trained models, a log likelihood objective is minimized including the *Mahalanobis distance* as statistical measure. After training, the obtained model defines a 'style-space' with item embeddings where items which are related define a 'style' in a quantifiable manner using the Mahalanobis distance among their embeddings – in other words, this approach allows to predict (solely based on visual features) which items go together and which do not. The approach can be easily extended to also enable personalized (user specific) recommendations.

A conceptually similar approach is outlined in [13] where pairs of items belonging to different categories, e.g., shoes and pants, are taken as inputs to train a Siamese CNN. Just as in the case of [17], the resulting item embeddings give rise to a style space. The discussed image-based recommender models are metric-learning approaches that rely on the metric trained with the selected data. This type of solution depends on explicit similarity functions and category trees, which contain hidden assumptions that compromise granularity, to evaluate substitutes and complements. As a result, those methods might be not flexible enough to describe items with complex relationships.

To address this shortcoming, the Mixtures of Non-Metric Embeddings for Recommendation (MONOMER) was proposed [19]. The key idea of MONOMER is to treat the distance metric (by giving up some of its mathematical properties) as a weighted ensemble composed of expert models of various aspects of the visual style. The embedding of the target represents a reference vector which is evaluated against all the potential matches in the ensemble. Finally, the probability of the edge is obtained as a sum over all the expert models.

Besides traditional neural architectures that utilize them as inputs for binary classification, the Complementary Recommendation using Adversarial Feature Transform (CRAFT) [30] provides a new angle to consume the style space using the emerging Generative Adversarial Network (GAN) [31]. Briefly, a GAN is an unsupervised paradigm composed of a discriminator for binary classification and a generator that produces synthesized samples to fool the former. The discriminator and generator are trained simultaneously in an adversarial manner so the generator eventually becomes able to synthesize real samples. The core of CRAFT is a GAN that takes concatenated image embeddings of co-purchase pairs, which are treated as supervised signals, as input and generates complements in the style space. Finally, the generated complementary embeddings are fed into a nearest neighbour search with a pool of candidate products to select the top-N recommendations.

2) Textual model: Despite the success of visual recommenders in modeling complementary relationships, the high cost incurred during data acquisition, data processing, and model training usually limits the application of such solutions. In comparison, text features of products such as titles and descriptions are easy to utilize and it is assumed that they often contain explicit attributes that imply product relationships. Proceeding from this assumption, the deep style match model [32] represents items by their title embeddings and evaluates the pairing relationship by measuring style compatibility. Specifically, item embeddings are fed into a convolutional sentence model with multiple feature maps followed by a siamese convolutional model for binary classification. Compared with the image-based models [13], [17] that require expensive training to alleviate visual noise, the proposed text-based solution offers better accuracy and scalability.

Another text-based model is *sceptre* that stands for Substitute and Complementary Edges between Products from Topics in Reviews [15]. As the name implies, this model utilizes product reviews to glean substitute and complementary relationships. The objective of *sceptre* is a combination of topic modeling as described by Latent Dirichlet Allocation (LDA) [33] and edge detection of related topics which is modeled by logistic regression. In the training process, the update of the joint objective alternates between updating the topic vectors and updating the probabilities of directional edges between topics to converge towards an optimum. Additional features (apart from text features) are taken into account to help sharpen the directional relationship between two products such as the differences in price and star-rating. In order to constrain the topic space for training efficiency, topics are associated with their own category trees that are subsets of the full category hierarchy. Compared with image-based methods, *sceptre* can easily handle the cold-start problem by utilizing universal text sources.

3) Multimodal input model: Considering the individual advantages of visual and textual inputs, combining both of them becomes an intuitive direction to improve the model. With the help of neural networks, complex hidden characteristics and flexible input sources could be simultaneously utilized to formulate various relationships among users, items, and user-item pairs. In comparison with models that rely on a single type of input, the Neural Complementary Recommender (ENCORE) [14] is a model that jointly utilizes text, image, and user preference feature as input to classify the complementary relationship between two items. More specifically, the algorithm is composed of three steps: 1) detecting complements by measuring the embedding distances of text and image for two items, 2) augmenting the complementary distance with user preference, and 3) training the neural network with the embeddings distances and user preference concatenated as the input. Eventually, the co-purchase data of Amazon, which include ”Bought together” and ”Also bought” records, are adopted for training and evaluation.

4) Co-occurence embedding model: Feature-based models can yield item representations that contain rich information, however, the compute and storage demand may limit their scalability. Instead, the substitute and complementary relationship can be modeled using the co-occurrence of listings that is widely adopted in unsupervised solutions. To learn the embeddings from listing co-occurrence, the RSC framework [2] was proposed. The RSC framework aims to learn listing embeddings for both substitutes and complements by concurrently satisfying three objectives for the substitute network, complementary network, and user ratings. In detail, the co-views for substitutes, co-purchases for complements, and ratings are learned via pairwise ranking, skip gram, and matrix factorization, respectively. The resulting embeddings are applicable in both contexts of substitutional and complementary recommendations. Thus, RSC represents a promising attempt to combine supervised and unsupervised learning methods to enhance complementary recommendations.

Co-occurence of items contain category relation and multi-step links which turn out to be significant for improving accuracies in link predictions for both substitutional and complementary recommendations. To this end, the Path-constrained Method to discriminate Substitutes and Complements (PMSC) [18] is proposed. The initial step of PMSC is to determine the directed edge between two items based on the idea of Item2vec [34]. This is followed by a latent vector learning that recognizes the type of relation by projecting item embeddings to a relation space. The most significant component of PMSC deals with path constraints that model category-level mapping and multi-hop relation where item pairs are converted to category pairs to alleviate sparsity for the cold start situation. In addition, item relationships are extended to two hops that include three items as three pairs. Finally, the above constraints are jointly trained. With the help of the more sophisticated graph, PMSC significantly outperforms *spectre*.

The co-occurence of items could also be represented by exponential family embeddings (*ef-emb*) [16]. It generalizes the idea of word embeddings, which was put forward in the continuous bag-of-words framework of Word2Vec [35], to be applicable also to other high-dimensional data. In the context of shopping scenarios, *ef-emb* methods seek to represent the target item by it’s co-purchased items. The corresponding conditional probability is modeled by a distribution from the exponential family (e.g. Poission) that allows to linearly combine item embeddings and the corresponding user embeddings which are sampled from the purchase matrix. As it was shown in [16], in addition to market basket analysis, *ef-emb* can also be applied in other contexts such as neural science and text modeling. Cold-start users and items are not addressed in this framework, however, as the generation of embeddings relies intrinsically on interactions.

5) Sequence model: In most cases, substitute and complementary recommenders are functionally different. However, such a separation is not strictly possible where the search behavior of users can rapidly change during sessions. Such situations would need to be addressed by a model that is able to dynamically switch between those two modes. For this purpose, the Contextual Recurrent Neural Networks (CRNN) [20] was proposed. The CRNN is composed of multiple Gated Recurrent Units (GRUs) that utilizes contextual features including timestamps, time gap between events, and event types besides the sequence of viewed listings. Those contextual information makes the model capable of predicting which kind of recommendations (substitutional or complementary) would be preferable to the user in each step of the session. Without the need to explicitly define substitutes and complements, CRNN automatically learns the relationships and provides an outlook to understand and solve this problem.

Considering the success of CRNN in capturing user activity and the advantage of multimodal-input style modeling in fashion recommendation [14], the Bidirectional Long Short Term Memory (Bi-LSTM) [36] was proposed to exploit the merits of both these concepts. In this work, image embeddings are fed into a Bi-LSTM to capture the sequential connection of co-purchases in visual space. During training, the image and text embeddings are jointly trained so as to project them into the same visual-semantic embedding space. The final solution consists of multiple tasks including blank filling, outfit generation, and compatibility prediction.

TABLE I
DATASET CHARACTERISTICS

	Users/Sessions	Items	Density (%)	Scale	N Items/User μ	N Items/User σ
Amazon (Cell Phones)	2261045	319678	0.0005	1–5	1.5	1.5
Amazon (Clothing and Fashion)	3117268	1136004	0.0002	1–5	1.8	2.5
Amazon (Electronics)	4201696	476002	0.0004	1–5	1.9	2.9
Amazon (Kitchen)	2511610	410243	0.0004	1–5	1.7	2.3
Amazon (Office Products)	909314	130006	0.001	1–5	1.4	1.5
Movie Lens Large (20M)	138493	26744	0.54	0.5–5	144.	230.
Movie Lens Small (100k)	610	9724	1.70	0.5–5	165.	269.
Polyvore	17316	72413	0.009	N/A	6.6	1.4
RecSys/CIKM	87934	71799	0.005	N/A	3.6	4.2
YooChoose	9249729	52739	0.005	N/A	2.9	3.0

IV. Complementary recommender Evaluation

Performance evaluation for the existing solutions can be divided into two main approaches: online and offline evaluations. Whilst online validation may offer a good insight into real-world performance, such results are virtually impossible to replicate. On the other hand, with offline validation, we are able to further characterize the data sets and metrics chosen by the literature discussed.

A. Evaluation Data Sets

The characteristics of the off-line data sets used by the existing solutions are summarized in TABLE I. The most striking feature of these data sets is that they are often highly sparse, with densities of the order 10^{-5}–10^{-4} %, where we defined density by $\#ratings/(\#items \cdot \#users) \times 100$.

The Movie Lens [16] data set is an outlier in this regard, having a density of 0.54% and a very high ratio $\#\text{items}/\#\text{users}$ given by $\mu = 144$. Furthermore, whilst most data sets adopt the *user-id*, *item-id*, *rating* schema, there are significant variations. The Polyvore data set consists of fashion outfits, each containing multiple fashion items. The YooChoose [20] data set on the other hand, records *session-id* rather than *user-id* where each session contains multiple views. Some data sets deal with implicit ratings, such as Polyvore [36] and YooChoose, while others adopt explicit ratings. In all cases the star-rating ranges from 0.5 to 5.

B. Evaluation Metrics

A brief description of some of the recommendations specific metrics used by the surveyed papers is given below.

1) Precision @ k: Adopted by solutions such as PMSC [18] and Sceptre [15], this metric captures the ability of a recommender system to predict accurately the items from the test set the user has interacted with. It is given as $\#\text{true\_positives}/k$, where k is the number of the top k recommendations. Typically a single interaction is held out for each user, in which case it equates to the hit rate @ k.

2) ROC Curve: Veit et. al. [13] calculate the roc curve by sweeping a threshold value and calculating the true positive and false positive rates respectively. This technique allows the observation of the models performance at a number of threshold points, helping select a good threshold as well as offering a deeper understanding of model performance.

3) AUC Score: An AUC (Area Under The Curve) score can be calculated by integrating the area under the ROC curve, as is done by Han et. al. [36].

4) FITB Accuracy: "Fill in the blank" accuracy was used by Han et. al. [36]. Here, for each outfit in the test set, a single item was removed. After each missing item is predicted FITB accuracy can be calculated as: $\#\text{true\_positives}/\#\text{total\_test\_outfits}$.

V. Conclusions

Recommender systems, which can be classified as substitute and complementary types, are one of the most effective solutions to boost business revenue. Therefore, the research and development of this topic have been rapidly progressing in recent years. Compared with the majority of the research that focuses on substitute recommenders, this survey reviews the currently available solutions and provides specific insights for complementary recommendations in three aspects including identifying the ground truth for complements, modeling complementary relationships, and evaluating complementary recommenders. The information aggregated in our work could provide researchers with a solid reference that enables future advancement for this field.

References

[1] K. Hosanagar, D. Fleder, D. Lee, and A. Buja, "Will the global village fracture into tribes? recommender systems and their effects on consumer fragmentation," *Management Science*, vol. 60, no. 4, pp. 805–823, 2013.

[2] T. Zhao, J. McAuley, M. Li, and I. King, "Improving recommendation accuracy using networks of substitutable and complementary products," in *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 3649–3655.

[3] M. Zhang and J. Bockstedt, "Complements and substitutes in product recommendations: The differential effects on consumers' willingness-to-pay," *CEUR Workshop Proceedings*, vol. 1679, pp. 36–43, 2016.

[4] J. Bobadilla, F. Ortega, A. Hernando, and A. GutiÃ©rrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109 – 132, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705113001044

[5] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: a literature survey," *International Journal on Digital Libraries*, vol. 17, no. 4, pp. 305–338, Nov 2016. [Online]. Available: https://doi.org/10.1007/s00799-015-0156-0

[6] M. Kunaver and T. PoÅ¾rl, "Diversity in recommender systems â€" a survey," *Knowledge-Based Systems*, vol. 123, pp. 154 – 162, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705117300680

[7] Z. Batmaz, A. Yurekli, A. Bilge, and C. Kaleli, "A review on deep learning for recommender systems: challenges and remedies," *Artificial Intelligence Review*, Aug 2018. [Online]. Available: https://doi.org/10.1007/s10462-018-9654-y

[8] J. Bentahar, M. Taghavi, K. Bakhtiyari, and C. Hanachi, "New Insights Towards Developing Recommender Systems," *The Computer Journal*, vol. 61, no. 3, pp. 319–348, 06 2017.

[9] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," *Expert Systems with Applications*, vol. 97, pp. 205 – 227, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417417308333

[10] J. Zheng, X. Wu, J. Niu, and A. Bolivar, "Substitutes or complements: Another step forward in recommendations," in *Proceedings of the 10th ACM Conference on Electronic Commerce*, ser. EC '09. New York, NY, USA: ACM, 2009, pp. 139–146. [Online]. Available: http://doi.acm.org/10.1145/1566374.1566394

[11] T. Raeder and N. V. Chawla, "Market basket analysis with networks," *Social Network Analysis and Mining*, vol. 1, no. 2, pp. 97–113, Apr 2011. [Online]. Available: https://doi.org/10.1007/s13278-010-0003-7

[12] I. Trofimov, "Inferring complementary products from baskets and browsing sessions," *CoRR*, vol. abs/1809.09621, 2018.

[13] A. Veit, B. Kovacs, S. Bell, J. McAuley, K. Bala, and S. Belongie, "Learning visual clothing style with heterogeneous dyadic co-occurrences," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. ICCV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 4642–4650. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2015.527

[14] Y. Zhang, H. Lu, W. Niu, and J. Caverlee, "Quality-aware neural complementary item recommendation," in *Proceedings of the 12th ACM Conference on Recommender Systems*, ser. RecSys '18. New York, NY, USA: ACM, 2018, pp. 77–85. [Online]. Available: http://doi.acm.org/10.1145/3240323.3240368

[15] J. McAuley, R. Pandey, and J. Leskovec, "Inferring networks of substitutable and complementary products," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '15. New York, NY, USA: ACM, 2015, pp. 785–794. [Online]. Available: http://doi.acm.org/10.1145/2783258.2783381

[16] M. Rudolph, F. Ruiz, S. Mandt, and D. Blei, "Exponential family embeddings," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 478–486. [Online]. Available: http://papers.nips.cc/paper/6571-exponential-family-embeddings.pdf

[17] J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Image-based recommendations on styles and substitutes," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '15. New York, NY, USA: ACM, 2015, pp. 43–52. [Online]. Available: http://doi.acm.org/10.1145/2766462.2767755

[18] Z. Wang, Z. Jiang, Z. Ren, J. Tang, and D. Yin, "A path-constrained framework for discriminating substitutable and complementary products in e-commerce," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ser. WSDM '18. New York, NY, USA: ACM, 2018, pp. 619–627. [Online]. Available: http://doi.acm.org/10.1145/3159652.3159710

[19] R. He, C. Packer, and J. McAuley, "Learning compatibility across categories for heterogeneous item recommendation," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, Dec 2016, pp. 937–942.

[20] E. Smirnova and F. Vasile, "Contextual sequence modeling for recommendation with recurrent neural networks," in *Proceedings of the 2Nd Workshop on Deep Learning for Recommender Systems*, ser. DLRS 2017. New York, NY, USA: ACM, 2017, pp. 2–9. [Online]. Available: http://doi.acm.org/10.1145/3125486.3125488

[22] M. J. Zaki, "Generating non-redundant association rules," in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '00. New York, NY, USA: ACM, 2000, pp. 34–43. [Online]. Available: http://doi.acm.org/10.1145/347090.347101

[21] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proceedings of the 20th International Conference on Very Large Data Bases*, ser. VLDB '94. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994, pp. 487–499. [Online]. Available: http://dl.acm.org/citation.cfm?id=645920.672836

[23] K. Gouda and M. J. Zaki, "Efficiently mining maximal frequent itemsets," in *Proceedings 2001 IEEE International Conference on Data Mining*, Nov 2001, pp. 163–170.

[24] P.-N. Tan, V. Kumar, and J. Srivastava, "Selecting the right objective measure for association analysis," *Information Systems*, vol. 29, no. 4, pp. 293 – 313, 2004, knowledge Discovery and Data Mining (KDD 2002). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306437903000723

[25] G. Pandey, S. Chawla, S. Poon, B. Arunasalam, and J. G. Davis, "Association rules network: Definition and applications," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 1, no. 4, pp. 260–279. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.10027

[26] H. Hwangbo, Y. S. Kim, and K. J. Cha, "Recommendation system development for fashion retail e-commerce," *Electronic Commerce Research and Applications*, vol. 28, pp. 94 – 101, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1567422318300152

[27] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '00. New York, NY, USA: ACM, 2000, pp. 1–12. [Online]. Available: http://doi.acm.org/10.1145/342009.335372

[28] H. Zhao, L. Si, X. Li, and Q. Zhang, "Recommending complementary products in e-commerce push notifications with a mixture model approach," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '17. New York, NY, USA: ACM, 2017, pp. 909–912. [Online]. Available: http://doi.acm.org/10.1145/3077136.3080676

[29] M. Grbovic, V. Radosavljevic, N. Djuric, N. Bhamidipati, J. Savla, V. Bhagwan, and D. Sharp, "E-commerce in your inbox: Product recommendations at scale," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '15. New York, NY, USA: ACM, 2015, pp. 1809–1818. [Online]. Available: http://doi.acm.org/10.1145/2783258.2788627

[30] C. P. Huynh, A. Ciptadi, A. Tyagi, and A. Agrawal, "Craft: Complementary recommendation by adversarial feature transform," in *Computer Vision – ECCV 2018 Workshops*, L. Leal-Taixé and S. Roth, Eds. Cham: Springer International Publishing, 2019, pp. 54–66.

[31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[32] K. Zhao, X. Hu, J. Bu, and C. Wang, "Deep style match for complementary recommendation," in *AAAI Workshops*, 2017. [Online]. Available: https://aaai.org/ocs/index.php/WS/AAAIW17/paper/view/15069

[33] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[34] O. Barkan and N. Koenigstein, "Item2vec: Neural item embedding for collaborative filtering," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, Sep. 2016, pp. 1–6.

[35] G. C. Tomas Mikolov, Kai Chen and J. Dean, "Efficient estimation of word representations in vector space," in *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013.

[36] X. Han, Z. Wu, Y.-G. Jiang, and L. S. Davis, "Learning fashion compatibility with bidirectional lstms," in *Proceedings of the 25th ACM International Conference on Multimedia*, ser. MM '17. New York, NY, USA: ACM, 2017, pp. 1078–1086. [Online]. Available: http://doi.acm.org/10.1145/3123266.3123394