

# Personalized Entity Resolution with Dynamic Heterogeneous Knowledge Graph Representations

Ying Lin\*, Han Wang, Jiangning Chen, Tong Wang, Yue Liu,  
Heng Ji, Yang Liu, Premkumar Natarajan

Amazon Alexa AI

{linzying, wngn, cjiangni, tonwng, jihj, yangliud, premkntat}@amazon.com

## Abstract

The growing popularity of Virtual Assistants poses new challenges for Entity Resolution, the task of linking mentions in text to their referent entities in a knowledge base. Specifically, in the *shopping* domain, customers tend to mention the entities implicitly (e.g., “organic milk”) rather than use the entity names explicitly, leading to a large number of candidate products. Meanwhile, for the same query, different customers may expect different results. For example, with “add *milk* to my cart”, a customer may refer to a certain product from his/her favorite brand, while some customers may want to re-order products they regularly purchase. Moreover, new customers may lack persistent shopping history, which requires us to enrich the connections between customers through products and their attributes. To address these issues, we propose a new framework that leverages personalized features to improve the accuracy of product ranking. We first build a cross-source heterogeneous knowledge graph from customer purchase history and product knowledge graph to jointly learn customer and product embeddings. After that, we incorporate product, customer, and history representations into a neural reranking model to predict which candidate is most likely to be purchased by a specific customer. Experiment results show that our model substantially improves the accuracy of the top ranked candidates by 24.6% compared to the state-of-the-art product search model.

## 1 Introduction

Given an entity mention as a query, the goal of entity resolution (or entity linking) (Ji and Grishman, 2011) is to link the mention to its corresponding entry in a target knowledge base (KB). In an academic shared task setting, an entity mention

is usually a name string, which can be a person, organization or geo-political entity in a news context, and the KB is usually a Wikipedia dump with rich structured properties and unstructured text descriptions. State-of-the-art entity resolution methods can achieve higher than 90% accuracy in such settings (Ji and Grishman, 2011; Ji et al., 2015; Agarwal and Bikel, 2020), and they have been successfully applied in hundreds of languages (Pan et al., 2017) and various domains such as disaster management (Zhang et al., 2018a) and scientific discovery (Zheng et al., 2014; Wang et al., 2015). Therefore, we tend to think entity resolution is a solved problem in academia. However, in industry, with the rise in popularity of Virtual Assistants (VAs) in recent years, an increasing number of customers now rely on VAs to perform daily tasks involving entities, including shopping, playing music or movies, calling a person, booking a flight, and managing schedules. The scale and complexity of industrial applications presents the following unique new challenges.

**Unpopular majority.** There is a massive number of new entities emerging every day. The entity resolver may know very little about them since very few users interact with them. Handling these tail entities effectively requires the use of property linkages between entities and shared user interests. Similarly, there might be many new users with limited interaction history, and we need to infer their interests from other users who have interacted with similar entities.

**Large number of ambiguous variants.** When interacting with VAs, users tend to use short and less informative utterances with the expectation that the VAs can intelligently infer their actual intentions. This raises the need for *personalization* when resolving the entities. In the *shopping* domain, this problem is even more challenging as customers typically use implicit entity mentions

\* This work was done when the first author was on an internship at Amazon Alexa AI.

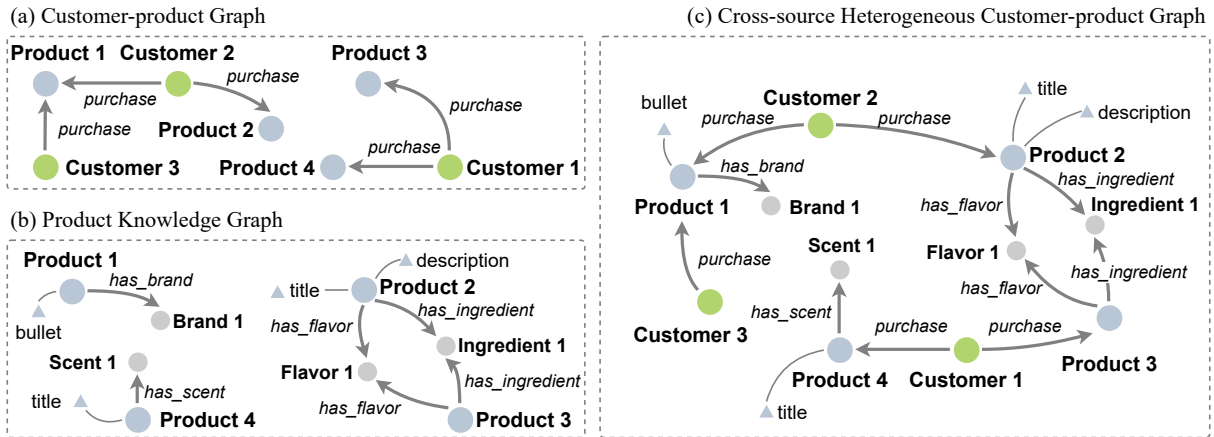


Figure 1: An illustration of the cross-source heterogeneous customer-product graph.

(e.g., “organic milk”) instead of explicit names (e.g., “Horizon Organic Shelf-Stable 1% Lowfat Milk”) which usually leads to a large number of candidates due to the ambiguity. However, with VAs’ voice user interface (VUI), the number of products that can be presented to the customers is very limited, if not only one. In this work, we focus on the problem of *personalized entity resolution* in the shopping domain. Given a query and a list of retrieved candidates, we aim to return the product that is most likely to be purchased by a customer.

**Beyond ambiguity.** In the traditional news entity linking setting, each entity in the KB refers to a unique world object. In contrast, in e-commerce, the same product can have multiple variants. For example, a customer may like to stick to a toothpaste product of a certain brand and flavor, but choose different sizes (thus different entities) in each purchase. These entities in the target KB refer to the same product but have different properties (in this case, size). Therefore it is important to construct fine-grained knowledge graphs to profile products and capture the implicit connections between customers based on the properties of their purchased products.

We make three assumptions: (H1) customers tend to purchase products they have purchased in the past; (H2) customers tend to purchase products that share some properties; (H3) customers who purchased products with similar properties share similar interests. Based on these assumptions, we propose to represent customers and products as low-dimensional vectors learned from a graph of customers and products. Unlike social media networks with rich interactions among users, the customers of most shopping services are isolated,

which prevents us from learning customer embeddings effectively. To address this issue, we propose to build a cross-source heterogeneous knowledge graph as Figure 1 depicts to indirectly establish rich connections among customers from a) users’ purchase history (customer-product graph) and b) a product knowledge graph, and further jointly learn the representations of nodes in this graph using a Graph Neural Network (GNN)-based method. In Figure 1(c), for instance, we can build connections between Customer 1 and Customer 2 because their purchased products share the same ingredient attribute, and thus possibly recommend Product 2 to Customer 1 even though it does not appear in his/her purchase history. In addition to static customer embeddings, we further propose an attentive model to dynamically generate a history representation for each user based on the current query. Finally, the model predicts how likely a candidate will be purchased using entity, customer, and history representations.

Experiments on real purchase records collected from an online shopping service show that our method significantly improves the purchase rate of the top ranked products.

## 2 Methodology

Given a query  $q$  from a customer  $c$ , and a list of candidate products  $\mathcal{P} = \{p_1, \dots, p_L\}$ , where  $L$  is the number of candidates, our goal is to predict the product that the customer will purchase based on their purchase history and the product knowledge graph. Specifically, we use purchase records  $\{r_1, \dots, r_H\}$  where  $H$  is the number of historical records. As Figure 2 illustrates, we jointly learn customer and product embeddings from a cross-

source customer-product graph using GNN. To perform personalized ranking, we incorporate the learned customer embedding and history representation as additional features when calculating the confidence score of each candidate. We then rank all candidates by confidence score and return the top one.

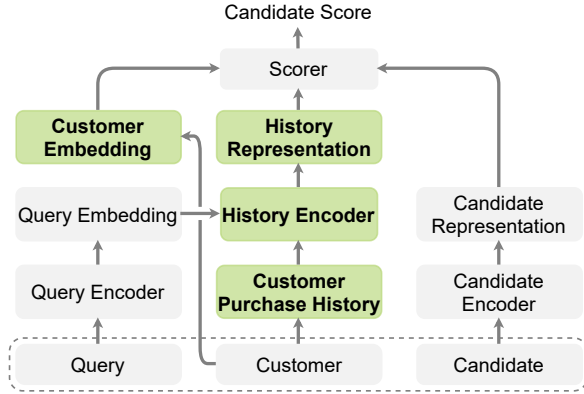


Figure 2: An illustration of our framework.

## 2.1 Candidate Retrieval

We first retrieve candidate products for each query using QUARTS (Nigam et al., 2019; Nguyen et al., 2020), which is an end-to-end neural model for product search. QUARTS has three major components: (1) an LSTM-based (long short-term memory) classifier adapted from the entailment model in (Rocktäschel et al., 2016) to predict whether a product-query pair is matched; (2) a variational encoder-decoder (VED) query generator that generates difficult negative examples to tackle the class imbalance issue in the training data as a search engine typically returns much more matched query-item examples than mismatched ones, and (3) a state combiner that switches between query representations computed by the classifier and generator. During training, the VED generator takes as input a matched product-query pair  $(I, Q)$  and generates a mismatched query  $Q_{\text{gen}}$  which is lexically similar to  $Q$ . The state combiner then merges  $H$ , the representation computed by the classifier, and  $H_{\text{gen}}$ , the representation computed by the generator, as  $sH_{\text{gen}} + (1 - s)H$ , where  $s$  is a binary value that controls which query to use and whether the gradients are back-propagated to the classifier or generator.

## 2.2 Joint Customer and Product Embedding

The next step is to obtain the representations of customers and products. Customer embeddings are usually learned from user-generated texts (Preoțiuc-Pietro et al., 2015; Yu et al., 2016; Ribeiro et al., 2018) or social relations (Perozzi et al., 2014a; Grover and Leskovec, 2016; Zhang et al., 2018b), neither of which are available in the shopping dataset we use. Alternatively, we establish *indirect* connections among customers through their purchased products under hypothesis H3, and form a customer-product graph as shown in Figure 1(a). This graph only contains a single type of relation (i.e., purchase) and ignores product attributes. As a result, it tends to be sparse and less effective for customer representation learning.

In order to learn more informative embeddings, we propose to incorporate richer information from a product knowledge graph (Figure 1(b)) where products are not only connected to different attribute nodes (e.g., brands, flavors), but they may also be associated with textual features (e.g., title) and boolean features (e.g., isOrganic, encoded as a boolean vector).

By merging the product knowledge graph and the customer-product graph, we obtain a more comprehensive graph (Figure 1(c)) of higher connectivity. For example, in the original customer-product graph, Customer 1 and Customer 2 are disconnected because they do not share any purchase. In the new graph, they have an indirect connection through Product 2 and Product 3, which share the same flavor and ingredient.

From this heterogeneous graph, we jointly learn customer and product representations using a two-layer Relational Graph Convolutional Network (Schlichtkrull et al., 2018). The embedding of each node is updated as:

$$\mathbf{h}_i^{l+1} = \text{ReLU}\left(\mathbf{W}_0^l \mathbf{h}_i^l + \sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{|N_i^r|} \mathbf{W}_r^l \mathbf{h}_j^l\right),$$

where  $\mathbf{h}_i^l$  is the representation of node  $i$  at the  $l$ -th layer,  $N_i^r$  is the set of neighbor indices of node  $i$  under relation  $r \in R$ , and  $\mathbf{W}_0^l$  and  $\mathbf{W}_r^l$  are learnable weight matrices.

In order to capture textual features (i.e., product titles, descriptions, and bullet<sup>1</sup>), we use a pre-trained RoBERTa (Liu et al., 2019) encoder to generate a fix-sized representation for each product.

<sup>1</sup>Bullet points that outline the main features of a product.

Specifically, we concatenate textual features using a special separator token [SEP], obtain the RoBERTa representation for each token, and then use the averaged embedding to represent the whole sequence. To reduce the runtime, we calculate customer and product embeddings offline and cache the results.

### 2.3 Candidate Representation

In addition to the product embedding, we further incorporate the following features to enrich the representation of each candidate.

**Rank:** the order of the candidate returned by the product retrieval system.

**Relative Price:** how much a product’s absolute price is higher or lower than the average price of all retrieved candidates as price is an important factor that affects purchasing decision.

**Previously Purchased:** a binary flag indicating whether a candidate has been purchased by the customer or not.

**Textual Feature:** we use RoBERTa to encode each candidate’s textual features (i.e., title, bullet, description). This RoBERTa encoder is fine-tuned during training.

We concatenate these features with the product embedding and project the vector into a lower dimensional space using a feed forward network.

### 2.4 History Representation

Although customer embeddings can encode purchase history information, they are *static* and may not effectively provide the most relevant information for each specific query. For example, if the query is “bookshelf”, the furniture-related purchase records are more likely to help the model predict the product that the customer will purchase, while if the query is “sulfate-free shampoo”, the purchase records of beauty products are more relevant. To tackle this issue, we propose to generate a *query-aware* history representation  $\mathbf{v}$  based on the current query  $\mathbf{q}$  from all purchase record representations  $\{\mathbf{v}_1, \dots, \mathbf{v}_H\}$  of the customer.

We first represent each purchase record as the concatenation of the product embedding, product price, and purchase timestamp. The query-aware history representation is then calculated as a weighted sum of the customer’s purchase record representations using an attention mechanism as follows.

$$e_i = \mathbf{v}^\top \tanh(\mathbf{W}_q \mathbf{q} + \mathbf{W}_v \mathbf{v}_i),$$

$$a_i = \text{Softmax}(e_i) = \frac{\exp(e_i)}{\sum_k^M \exp(e_k)},$$

$$\mathbf{v} = \sum_i^H a_i \mathbf{v}_i,$$

where  $\mathbf{v}^\top$ ,  $\mathbf{W}_q$ , and  $\mathbf{W}_v$  are learnable weights.

### 2.5 Candidate Ranking

We adopt a feed forward neural network that takes in the candidate, customer, and history representations, and returns a confidence score  $\hat{y}_i$  which indicates how likely a candidate will be purchased. The confidence score is scaled to  $(0, 1)$  using a Sigmoid function. During training, we optimize the model by minimizing the following binary cross entropy loss function.

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i),$$

where  $N$  denotes the total number of candidates, and  $y_i \in \{0, 1\}$  is the true label. In the inference phase, we calculate confidence scores for all candidates for each session and return the one with the highest score.

## 3 Experiment

### 3.1 Data

**Product Knowledge Graph.** In our experiment, we use a knowledge graph of products in five categories (i.e., grocery, beauty, luxury beauty, baby, and health care), which contains 24,287,337 unique product entities. As Figure 1(b) depicts, the products in this knowledge graph are connected through attribute nodes, including brands, scents, flavors, and ingredients. This knowledge graph also provides rich attributes for each product node. We use two types of attributes in this work, textual features (i.e., title, description, and bullet) and binary features (e.g., isOrganic, isNatural).

**Evaluation Dataset.** We randomly collect 1 million users’ purchase sessions from November 2018 to October 2019 on an online shopping service. Each session contains a query, an obfuscated identifier, a timestamp, and a list of candidate products retrieved using QUARTS where only one product is purchased.

We split the sessions before and after September 1, 2019 into two subsets. The first subset only

serves as the purchase history and is used to construct the customer-product graph. From the second subset, we randomly sample 22,000 customers with at least one purchase record in the first subset and take their last purchase sessions for training or evaluation. Specifically, we use 20,000 sessions for training, 1,000 for validation, and 1,000 for test. If a customer has multiple purchase sessions in the second subset, other sessions before the last one are also considered as purchase history when we generate history representations, while they are excluded from the customer-product graph, which is constructed from the first subset.

### 3.2 Experimental Setup

We optimize our model with AdamW (Loshchilov and Hutter, 2018) for 10 epochs with a learning rate of 1e-5 for the RoBERTa encoder, a learning rate of 1e-4 for other parameters, weight decay of 1e-3, a warmup rate of 10%, and a batch size of 100.

To encode textual features, we use the RoBERTa base model<sup>2</sup> with an output dropout rate of 0.5. To represent query words, we use 100-dimensional GloVe embeddings (Pennington et al., 2014) pre-trained on Wikipedia and Gigaword<sup>3</sup>. We set the size of pre-trained customer and product embeddings to 100 and freeze them during training.

We use separate fully connected layers to project candidate and history representations into 100-dimensional feature vectors before concatenating them for ranking. We use a two-layer feed forward neural network with a hidden layer size of 50 as the ranker and apply a dropout layer with a dropout rate of 0.5 to its input.

### 3.3 Quantitative Analysis

We compare our model to the state-of-the-art product search model QUARTS as the baseline. Because our target usage scenarios are VAs where only one result will be returned to the user, we use accuracy as our evaluation metric. We implement the following baseline ranking methods.

**Purchased:** We prioritize products previously purchased by the customer. If multiple candidates are previously purchased, we return the one ranked highest by QUARTS.

**Complex:** Customer and product embeddings are learned using Complex (Trouillon et al., 2016), a

<sup>2</sup>[https://huggingface.co/transformers/pretrained\\_models.html](https://huggingface.co/transformers/pretrained_models.html)

<sup>3</sup><https://nlp.stanford.edu/projects/glove/>

widely used knowledge embedding model that represents nodes in a knowledge graph as complex vectors and is able to capture antisymmetric relations using efficient dot product.

In Table 1, we show the relative gains compared to the baseline model QUARTS. With personalized features, our method effectively improves accuracy on both development and test sets.

We also conduct ablation studies by removing the following features and show results in Table 2.

**Ranking:** In this setting, our model ignores the original retrieval ranking returned by QUARTS.

**Personalized Features:** We remove personalized features (e.g., customer embedding, whether a product is previously purchased) in this setting.

**Product Embedding:** We remove pre-trained product embedding but still use textual features and binary features to represent products.

**Joint Embedding:** Customer and product embeddings are not jointly learned from the merged graph. Alternatively, customer embeddings are learned from the customer-product graph, and product embeddings are learned from the product knowledge graph.

In Table 2, from the results of Methods 6 and 7, we can see that removing either product or customer embedding degrades the performance of the model. The result of Method 8 shows that embeddings jointly learned from the merged cross-source graph achieve better performance on our downstream task. We also observe that the ranking returned by the product search system is still an important feature as Method 6 shows.

	Method	Dev Accuracy	Test Accuracy
1	QUARTS	0.0	0.0
2	Purchased	+10.5	+8.5
3	Complex	+25.7	+16.1
4	Our Model	<b>+32.9</b>	<b>+24.6</b>

Table 1: Relative gains compared to QUARTS. (%)

	Method	Dev Acc	Test Acc
4	Our Model	<b>+32.9</b>	<b>+24.6</b>
5	w/o Ranking	-17.1	-20.4
6	w/o Personalized Features	-10.5	-18.0
7	w/o Product Embedding	+25.2	+19.0
8	w/o Joint Embedding	+28.1	+20.4

Table 2: Ablation study. (% relative gains compared to QUARTS.)

Query	Candidates	History
#1 vitamin c serum	<ul style="list-style-type: none"> <li>* [3] instanatural vitamin c serum with hyaluronic acid &amp; vit e - natural &amp; organic anti wrinkle ...</li> <li>* [1] truskin vitamin c serum for face, topical facial serum with hyaluronic acid, vitamin e, 1 fl oz</li> <li>* [2] vitamin c serum for face - anti aging facial serum</li> <li>* [4] vitamin c serum plus 2% retinol, 3.5% niacinamide, 5% hyaluronic acid, 2% salicylic acid ...</li> </ul>	<ul style="list-style-type: none"> <li>* foundation makeup brush flat top kabuki for face - perfect for blending liquid, cream or flawless powder</li> <li>* women’s rogain 5% minoxidil foam for hair thinning and loss, topical treatment for women’s hair ...</li> <li>* vita liberata advanced organics fabulous self-tanning gradual lotion with marula oil, 6.76 fl oz</li> <li>* instanatural vitamin c serum with hyaluronic acid &amp; vit e - natural &amp; organic anti wrinkle reducer ...</li> </ul>
Our model promotes candidate 3 as this product was purchased by the customer.		
#2 toothpaste	<ul style="list-style-type: none"> <li>* [2] crest 3d white whitening toothpaste, radiant mint, 3.5oz, twin pack</li> <li>* [1] crest + scope complete whitening toothpaste, minty fresh, 5.4 oz, pack of 3</li> <li>* [3] pronamel gentle whitening enamel toothpaste for sensitive teeth, alpine breeze-4 ounces (pack of 3)</li> <li>* [4] colgate cavity protection toothpaste with fluoride - 6 ounce (pack of 6)</li> </ul>	<ul style="list-style-type: none"> <li>* crest 3d white toothpaste radiant mint (3 count of 4.1 oz tubes), 12.3 oz packaging may vary</li> <li>* skindinavia the makeup of countrrol finishing spray, 8 fluid ounce</li> <li>* crest 3d white toothpaste radiant mint (3 count of 4.1 oz tubes), 12.3 oz packaging may vary</li> <li>* nivea shea daily mointure body lotion - 48 hour moisture for dry skin - 16.9 fl. oz. pump bottle, ...</li> </ul>
Although the previously purchased item is no longer available, with entity embedding learned from the cross-source graph, our model successfully promotes the most similar product.		
#3 sun dried tomatoes	<ul style="list-style-type: none"> <li>* [3] 365 everyday value, <b>organic</b> sundried tomatoes in extra virgin olive oil, 8.5 oz</li> <li>* [1] 35 oz bella sun luci sun dried tomatoes julienne cut in olive oil (original version)</li> <li>* [2] julienne sun-dried tomatoes - 16oz bag (kosher)</li> <li>* [4] organic sun-dried tomatoes with sea salt, 8 ounces - salted, non-gmo, kosher, raw, vegan, ...</li> </ul>	<ul style="list-style-type: none"> <li>* #1 usda <b>organic</b> aloe vera gel - no preservatives, no alcohol - from freshly cut usa grown 100% pure ...</li> <li>* <b>organic</b> aloe vera gel with 100% pure aloe from freshly cut aloe plant, not powder - no xanthan ...</li> <li>* <b>wicked joe organic</b> coffee wicked italian ground</li> <li>* thayers alcohol-free original witch hazel facial toner with aloe vera formula, clear, 12oz</li> </ul>
Our model promotes an organic product as the customer probably prefers organic products based on the shopping records.		

Table 3: Positive examples in the data set. Candidates are listed in the order returned by our method. The number before each candidate is the original ranking returned by QUARTS. In the candidate column, we highlight the purchased products. In the history column, we highlight related records.

### 3.4 Qualitative Analysis

In Table 3 and Table 4, we show some positive and negative examples in the data set. From Table 3 we can see that multiple sources of evidence in the constructed heterogeneous knowledge graphs are complimentary and the combination of them successfully promotes various entities which match customers’ interests.

Table 4 shows examples where our model fails to return the correct item. In many cases, such as Example #4, the purchased product and the top ranked one only differ in packaging size. We also observe that sometimes customers may not repurchase a product even if it is in the candidate list.

To better understand the remaining errors, we randomly sample 100 examples where our model fails to predict the purchased items. As Figure 3 illustrates, we analyze these examples and classify the possible reasons into the following categories.

**Different size.** The predicted product and ground truth are the same product but differ in size. For example, while our model predicts “Lipton Herbal

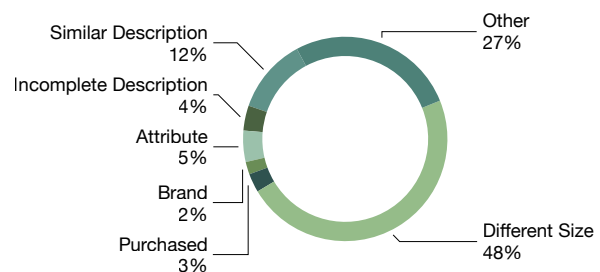


Figure 3: Distribution of remaining Errors.

Tea Bags, Peach Mango, 20 ct”, the customer purchases another item “Lipton Tea Herbal Peach Mango (pack of 2)”, which is actually the same product in 2 pack.

**Purchased.** The customer purchased the predicted product before but decides not to repurchase it. This usually happens in categories (e.g., toothpaste) where customers are more willing to try new products. Additionally, customers may be less likely to repurchase a product in some categories such as books and electronics.

**Uninformative title.** The purchased product has

Query	Candidates	History
#4 wasabi almonds	* [8] blue diamond almonds, bold wasabi & soy sauce, 16 ounce (pack of 1)	* epsom epsom salt 19 lb. bulk bag magnesium sulfate usp
	* [2] blue diamond almonds variety pack (1.5 ounce bags) (20 pack)	* blue diamond almonds, bold wasabi & soy sauce, 16 ounce (pack of 1)
	* [1] blue diamond almonds bold wasabi & soy sauce almonds, 25 ounce (pack of 1)	* signature trail mix, peanuts, m & m candies, raisins, almonds & cashews, 4 lb
	* [6] blue diamond almonds, bold wasabi & soy, 1.5 ounce (pack of 12)	* amazon brand - happy belly nuts, chocolate & dried fruit trail mix, 48 ounce
	Our model promotes candidate 8 which is previously purchased, whereas the customer selects another size.	
#5 cacao powder	* [5] anthony’s organic cocoa powder, 2 lb, batch tested and verified gluten free & non gmo	* anthony’s organic cocoa powder, 2 lb, batch tested and verified gluten free & non gmo
	* [1] viva naturals #1 best selling certified organic cacao powder from superior criollo beans, 1 lb bag	* vör all natural keto nut butter spread (10oz)   only two ingredients   no sugar, no salt   vegan ...
	* [2] navitas organics cacao powder, 16oz. bag - organic, non-gmo, fair trade, gluten-free	* anthony’s organic cocoa powder, 2 lb, batch tested and verified gluten free & non gmo
	* [3] terrasoul superfoods raw organic cacao powder, 1 lb - raw   keto   vegan	* nutiva organic, neutral tasting, steam refined coconut oil from non-gmo, sustainably farmed coconuts ...
	* [4] viva naturals certified organic cacao powder (2lb) for smoothie, coffee and drink mixes	
	Our model promote “Anthony’s Organic Cocoa Powder” as it has been purchased twice by the customer.	

Table 4: Negative examples in the data set.

an uninformative title and is therefore not promoted. For example, when the customer searches for “masaman curry paste maesri”, our model promotes “Maesri Thai Masaman Curry - 4 oz (pack of 4)”, while the customer purchases “6 Can (4oz. Each) of Thai Green Red Yellow Curry Pastes Set”, which is also a Maesri product, but this key information is missing from its title.

**Similar title.** The title of the predicted product is similar to the titles of some purchased products in the customer’s history in a less important aspect. For example, the model promotes a “moisturizing” shave gel because the customer has purchased a “moisturizing” body wash, whereas the customer decides to purchase a product for “sensitive skin”.

**Brand.** The customer has purchased one or more products of the same brand.

**Attribute.** The customer has purchased one or more products with the same attribute (e.g., organic, keto, kosher).

**Other.** The model may fail to predict the purchased item in other uncategorized cases. For example, when a customer searches for “nail clippers” but has purchased only food in the past, the model is unlikely to utilize the history records to improve the ranking.

### 3.5 Remaining Challenges

Although our framework can improve the accuracy of predicting products that will be purchased, there are still some remaining challenges. We pro-

pose the following potential solutions for further improvement.

**Incorporating more informative features.** Some important features that affect purchase decisions are still missing in our framework, such as the average rating, customer review comments, and number of ratings. For example, we may promote the highest rated product for a customer who usually buys products with high ratings.

**Building a more comprehensive cross-source customer-product graph.** In this work, we merge the customer-product graph and product knowledge graph into a single graph, which has been proved to produce better embeddings for our target task. A natural extension is to include records from more sources, such as music or video playing history, and multimedia features such as product pictures.

**Modeling the interactions among purchase behaviors.** Our current attention-based method that generates history representations is “flat” and ignores the relationship among purchase behaviors. For example, for a customer who previously purchases a pod coffee maker, we should promote coffee capsules in the candidates over coffee beans or grounds.

**Incorporating cohort features.** When dealing with customers with limited or even no previous shopping records, a step forward is to cluster customers and produce cohort-based representations. In this way, customers can be better represented collectively through other similar customers, espe-

cially when we combine their interaction history from other domains and build a more comprehensive graph as suggested above.

To sum up, shopping is a complex behavior, the importance weights of various features may vary across types of products. For instance, customers may like to stick to the same brand for beauty products while changing the size, depending on their needs. In contrast, for clothing customers may care more about visual features rather than text descriptions, and for books customers rarely purchase the same book more than once. To tackle these remaining challenges, we aim to extend our framework to incorporate more multimedia features, extract knowledge from review comments, and present recommendation results in a more hierarchical way by clustering variants of the same product and presenting their different attributes.

## 4 Related Work

### 4.1 Neural Entity Linking

A variety of neural models (Gupta et al., 2017; Kolitsas et al., 2018; Cao et al., 2018; Sil et al., 2018; Gillick et al., 2019; Logeswaran et al., 2019; Wu et al., 2019; Agarwal and Bikel, 2020) have been applied to entity linking in recent years. Compared to traditional entity linking, our task is different in three aspects: (1) Our mentions are typically vague and occur in uninformative contexts, such as “add *toothpaste* to my cart”; (2) A mention may be reasonably linked to multiple entities, while only one of them is considered “correct” (purchased by the customer); (3) The ground truth for the same mention can be different for different customers.

### 4.2 Personalized Recommendation

A recommender system is an information filtering system that aims to suggest a list of items in which a user may be interested. Content-based filtering (Billsus and Pazzani, 2000; Aciar et al., 2007; Wang et al., 2018) and collaborative filtering (Shardanand and Maes, 1995; Konstan et al., 1997; Linden et al., 2003; Zhao and Shang, 2010) are two common approaches used in recommender systems. In recent years, researchers have also applied neural methods to improve the quality of recommendations (Xue et al., 2017; He et al., 2017; Wang et al., 2019a,b). Recommender systems usually rank items based on the user’s past behaviors (e.g., purchasing, browsing, rating) and current context (Linden et al., 2003; Smith and Linden, 2017),

whereas the results are not constrained by queries. Instead, our task requires a specific query and only returns the product that is most likely to be purchased from a list of relevant candidates.

### 4.3 Graph Embedding

Various methods have been proposed to learn low-dimensional vectors for nodes in knowledge graphs. Knowledge graph embedding methods, such as TransE (Bordes et al., 2013), DistMult (Yang et al., 2014), ComplEx (Trouillon et al., 2016), and RotatE (Sun et al., 2018), typically represent the head entity, relation, and tail entity in each triplet in the knowledge graph as vectors and aim to rank true triplets higher than corresponding corrupted triplets. Matrix Factorization-based methods (He and Niyogi, 2004; Nickel et al., 2011; Qiu et al., 2018) represent the graph as a matrix and obtain node vectors by factorizing this matrix. Another category of frameworks (Perozzi et al., 2014b; Yang et al., 2015; Grover and Leskovec, 2016) use random walk to sample paths from the input graph and learn node embeddings from the sampled paths using neural models such as SkipGram and LSTM.

### 4.4 Heterogeneous Network

The earliest study of mining heterogeneous network dates back to (Sun et al., 2009), which coins the concept of Heterogeneous Information Network. After that, heterogeneous network has been applied to a range of tasks, including ranking (Ji et al., 2011), similarity search (Sun et al., 2011), link prediction (Dong et al., 2015), academic paper recommendation (Pan et al., 2015), and malicious account detection (Liu et al., 2018).

Recently, with the advent of graph neural network, many methods based on this new paradigm have been proposed to learn graph representations on heterogeneous graphs, such as Heterogeneous Graph Neural Network (HetGNN) (Zhang et al., 2019), Heterogeneous Graph Attention Network (HAN) (Wang et al., 2019c), and Heterogeneous Graph Transformer (HGT) (Hu et al., 2020).

## 5 Conclusion and Future Work

We propose a novel framework to jointly learn customer and product representations based on a cross-source heterogeneous graph constructed from customers’ purchase history and the product knowledge graph to improve personalized entity resolution. Experiments show that our framework can

effectively increase the purchase rate of the top ranked products. In the future, we plan to investigate better approaches to integrating personalized features and extend the framework to cross-lingual cross-media settings and generate conversations for more proactive and explainable entity recommendation and summarization.

## References

- Silvana Aciar, Debbie Zhang, Simeon Simoff, and John Debenham. 2007. Informed recommender: Basing recommendations on consumer product reviews. *IEEE Intelligent systems*, 22(3):39–47.
- Oshin Agarwal and Daniel M Bikel. 2020. Entity linking via dual and cross-attention encoders. *arXiv preprint arXiv:2004.03555*.
- Daniel Billsus and Michael J Pazzani. 2000. User modeling for adaptive news access. *User modeling and user-adapted interaction*, 10(2-3):147–180.
- Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26:2787–2795.
- Yixin Cao, Lei Hou, Juanzi Li, and Zhiyuan Liu. 2018. Neural collective entity linking. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 675–686.
- Yuxiao Dong, Jing Zhang, Jie Tang, Nitesh V Chawla, and Bai Wang. 2015. Coupledlp: Link prediction in coupled networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 199–208.
- Dan Gillick, Sayali Kulkarni, Larry Lansing, Alessandro Presta, Jason Baldrige, Eugene Ie, and Diego Garcia-Olano. 2019. Learning dense representations for entity retrieval. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 528–537.
- Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864.
- Nitish Gupta, Sameer Singh, and Dan Roth. 2017. Entity linking via joint encoding of types, descriptions, and context. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2681–2690.
- Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182.
- Xiaofei He and Partha Niyogi. 2004. Locality preserving projections. In *Advances in neural information processing systems*, pages 153–160.
- Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. 2020. Heterogeneous graph transformer. In *Proceedings of The Web Conference 2020*, pages 2704–2710.
- Heng Ji and Ralph Grishman. 2011. Knowledge base population: Successful approaches and challenges. In *Proc. ACL2011*.
- Heng Ji, Joel Nothman, Ben Hachey, and Radu Florian. 2015. Overview of tac-kbp2015 tri-lingual entity discovery and linking. In *Proc. Text Analysis Conference (TAC2015)*.
- Ming Ji, Jiawei Han, and Marina Danilevsky. 2011. Ranking-based classification of heterogeneous information networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1298–1306.
- Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. 2018. End-to-end neural entity linking. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 519–529.
- Joseph A Konstan, Bradley N Miller, David Maltz, Jonathan L Herlocker, Lee R Gordon, and John Riedl. 1997. Grouplens: applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3):77–87.
- Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Ziqi Liu, Chaochao Chen, Xinxing Yang, Jun Zhou, Xiaolong Li, and Le Song. 2018. Heterogeneous graph neural networks for malicious account detection. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 2077–2085.
- Lajanugen Logeswaran, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, Jacob Devlin, and Honglak Lee. 2019. Zero-shot entity linking by reading entity descriptions. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3449–3460.
- Ilya Loshchilov and Frank Hutter. 2018. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

- Thanh Nguyen, Nikhil Rao, and Karthik Subbian. 2020. [Learning robust models for e-commerce product search](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6861–6869, Online. Association for Computational Linguistics.
- Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *ICML*, volume 11, pages 809–816.
- Priyanka Nigam, Yiwei Song, Vijai Mohan, Vihan Lakshman, Weitian (Allen) Ding, Ankit Shingavi, Choon Hui Teo, Hao Gu, and Bing Yin. 2019. [Semantic product search](#). In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 2876–2885, New York, NY, USA. Association for Computing Machinery.
- Linlin Pan, Xinyu Dai, Shujian Huang, and Jiajun Chen. 2015. Academic paper recommendation based on heterogeneous graph. In *Chinese computational linguistics and natural language processing based on naturally annotated big data*, pages 381–392. Springer.
- Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In *Proc. the 55th Annual Meeting of the Association for Computational Linguistics (ACL2017)*.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. [Glove: Global vectors for word representation](#). In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- Bryan Perozzi, Rami Al-Rfou, and S. Skiena. 2014a. Deepwalk: online learning of social representations. In *KDD '14*.
- Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014b. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710.
- Daniel Preoțiuc-Pietro, Vasileios Lampsos, and Nikolaos Aletras. 2015. An analysis of the user occupational class through twitter content. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*.
- Jiezhong Qiu, Yuxiao Dong, Hao Ma, Jian Li, Kuansan Wang, and Jie Tang. 2018. Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 459–467.
- M. H. Ribeiro, Pedro H. Calais, Yuri A. Santos, V. Almeida, and W. Meira. 2018. Characterizing and detecting hateful users on Twitter. In *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)*.
- Tim Rocktäschel, Edward Grefenstette, K. Hermann, Tomáš Kociský, and P. Blunsom. 2016. Reasoning about entailment with neural attention. *CoRR*, abs/1509.06664.
- Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, pages 593–607. Springer.
- Upendra Shardanand and Pattie Maes. 1995. Social information filtering: algorithms for automating “word of mouth”. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210–217.
- Avirup Sil, Gourab Kundu, Radu Florian, and Wael Hamza. 2018. Neural cross-lingual entity linking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.
- Brent Smith and Greg Linden. 2017. Two decades of recommender systems at amazon.com. *Ieee internet computing*, 21(3):12–18.
- Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment*, 4(11):992–1003.
- Yizhou Sun, Jiawei Han, Peixiang Zhao, Zhijun Yin, Hong Cheng, and Tianyi Wu. 2009. Rankclus: integrating clustering with ranking for heterogeneous information network analysis. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, pages 565–576.
- Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2018. RotatE: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.
- Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. *International Conference on Machine Learning (ICML)*.
- Donghui Wang, Yanchun Liang, Dong Xu, Xiaoyue Feng, and Renchu Guan. 2018. A content-based recommender system for computer science publications. *Knowledge-Based Systems*, 157:1–9.
- Han Wang, Jin Guang Zheng, Xiaogang Ma, Peter Fox, and Heng Ji. 2015. Language and domain independent entity linking with quantified collective validation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 695–704.

- Hongwei Wang, Fuzheng Zhang, Mengdi Zhang, Jure Leskovec, Miao Zhao, Wenjie Li, and Zhongyuan Wang. 2019a. Knowledge-aware graph neural networks with label smoothness regularization for recommender systems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19*, page 968–977, New York, NY, USA. Association for Computing Machinery.
- Hongwei Wang, Fuzheng Zhang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2019b. Multi-task feature learning for knowledge graph enhanced recommendation. In *The World Wide Web Conference*, pages 2000–2010.
- Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019c. Heterogeneous graph attention network. In *The World Wide Web Conference*, pages 2022–2032.
- Ledell Wu, Fabio Petroni, Martin Josifoski, Sebastian Riedel, and Luke Zettlemoyer. 2019. Scalable zero-shot entity linking with dense entity retrieval. *arXiv preprint arXiv:1911.03814*.
- Hong-Jian Xue, Xinyu Dai, Jianbing Zhang, Shujian Huang, and Jiajun Chen. 2017. Deep matrix factorization models for recommender systems. In *IJCAI*, volume 17, pages 3203–3209. Melbourne, Australia.
- Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*.
- Zhilin Yang, Jie Tang, and William Cohen. 2015. Multi-modal bayesian embeddings for learning social knowledge graphs. *arXiv preprint arXiv:1508.00715*.
- Yang Yu, Xiaojun Wan, and Xinjie Zhou. 2016. User embedding for scholarly microblog recommendation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*.
- Boliang Zhang, Ying Lin, Xiaoman Pan, Di Lu, Jonathan May, Kevin Knight, and Heng Ji. 2018a. ELISA-EDL: A cross-lingual entity extraction, linking and localization system. In *Proc. The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT2018)*.
- Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. 2019. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 793–803.
- Zhen Zhang, Hongxia Yang, Jiajun Bu, Sheng Zhou, Pinggang Yu, Jianwei Zhang, Martin Ester, and Can Wang. 2018b. Anrl: Attributed network representation learning via deep neural networks. In *IJCAI*, volume 18.
- Zhi-Dan Zhao and Ming-Sheng Shang. 2010. User-based collaborative-filtering recommendation algorithms on hadoop. In *2010 Third International Conference on Knowledge Discovery and Data Mining*, pages 478–481. IEEE.
- Jin Guang Zheng, Daniel Howsmon, Boliang Zhang, Juergen Hahn, Deborah McGuinness, James Hendler, and Heng Ji. 2014. Entity linking for biomedical literature. In *BMC Medical Informatics and Decision Making*.