Prod2Vec - Distributed Semantic Representations of Retail Products based on Large-scale Transaction Logs

August 20, 2017

1 Problem Description

Co-occurrences of various products in retail transactions contain immense information about the competitive relationships of the products. Typical recommender system uses behavioral-based association measures to recommend various competitive products to the users. We propose a method to obtain vector representations of each product from a product universe which are optimized to capture the semantics of the products' appearances. The newer representation can be efficiently used in competitive product recommendation. The representations thus obtained are capable of showing semantic relationship among themselves. We present the task analogy where product vectors show meaningful relationships among themselves when vector arithmetic is performed. Furthermore, the vectors are also able to capture latent signals, which were implicit in the transactions for e.g. brand, category etc.

We have processed 18 million transactions consisting of unique 325,548 products to obtain vector representations. This also includes 1,551 categories. This is first of a kind of attempt to develop distributed representations of retail products from large-scale transaction data. Similar to the huge impact word vectors have in various applications of NLP, these product representations will aid to various applications like search, recommendation, identifying competitive product etc. The current process of obtaining these vectors is similar to GloVe in NLP which takes the co-occurrence matrix as the input. Log bilinear regression techniques like GloVe combines the advantages global matrix factorization and local context window methods to obtain the vector space representations of the tokens in question. We reformulate the problem statement as to map products onto a vector space where products with similar 'context' lie spatially closer to each other while contexually dissimilar items lie spatially farther apart.

2 Experiments

We have set the context window as 8, empirically for now (median basket size is 4). We also assume that a basket is a non-ordered collection of products, hence we chose to use asymmetric context window without distinguishing left and right context.

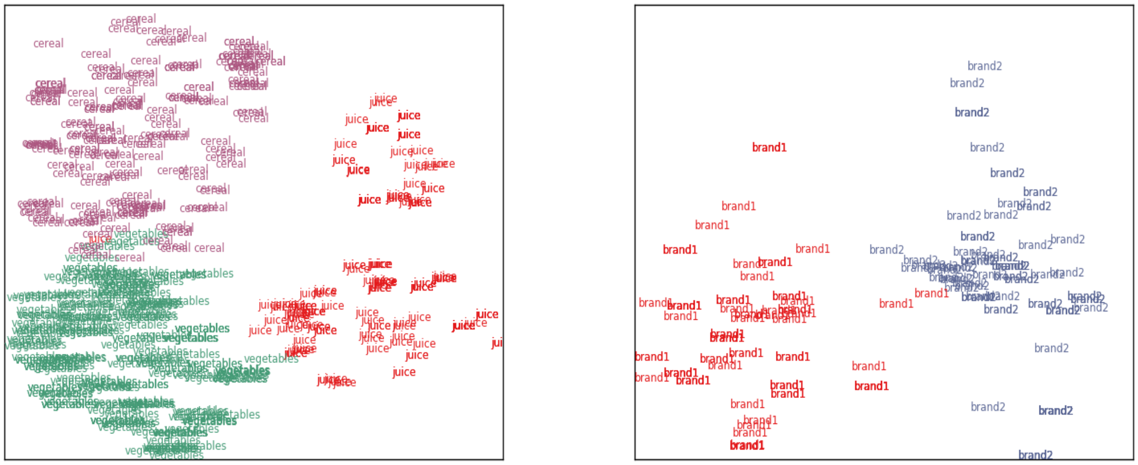
We also assumed that all baskets are independent. Hence, the context of a product should be delineated by the size of the basket it belongs. We introduce a basket of dummy products θ with size equal to the fixed context window between every basket so that the context window does not overlap two different basket transactions. As we penalizes tokens those occur rarely or too frequently and as the dummy product θ is going to appear too frequently, they will be automatically penalized without altering the embeddings generated from original products.

2.1 Analogy

In an NLP setting, *Analogy* is defined as “King is to Man as Queen is to?”. This is being done via vector arithmetic. Essentially, the analogous vector would the resultant vector of the operation (King - Man + Queen). Similarly, we experiment with product vectors from different categories. To illustrate, we choose three typical categories (Juice, Vegetable, Cereals) which are most frequently bought. Let’s assume Φ as the function which return the representation of the product argument. Then we obtain

$$\Phi(Juice1) + \Phi(Vegetable1) - \Phi(Juice2) = \Phi(Vegetable2)$$

We randomly sampled 1,000 products from the juice category and 1000 items from the vegetables category and retrieve the category of the resultant product. The highest **MRR** we obtain **0.472** along with highest **MAP@5** as **0.351**. This implies that out of 1,551 possible categories of items, the generated vector represents the required category 35% of the time. We also note that the remaining 65% is covered by 244 categories each contributing approximately 0.5% of the time. This can happen when customers like to buy most of their products in one visit and thus, there is a slight bit of noise with some products having no competitive relationships between them. Further to note, the optimal result we obtain with context window size of 8 and vector dimension of 200. We further plotted the product vectors those belong to category juice, vegetables and cereals. We also plot products from two different brands. Following are the visualizations -



(a) category clusters

(b) brand clusters

Figure 1: Two dimensional visualization of product vectors

It is important to note that, with having explicit incorporation of the product meta-data like brand, category etc., the product vectors, thus obtained, are naturally arranged in a semantically meaningful way in the new vector space. This strongly indicates that the transaction logs also capture the inherent product meta-information which drives consumers to buy.

2.2 Vector Arithmetic

We further extend the semantic analogy experiment to include brand names and product descriptions. Instead of only products, we augment the transaction logs by including brand and description tokens as separate entities. All unique brand and description tokens now have their corresponding co-occurrence scores. The model was trained with 325,548 such descriptions and 7,076 brands in addition to the individual products. With this inclusion, the number of vectors generated increased to 444,262. Table 1 shows some interesting illustrations. Since the vector representations generated are based on co-occurrence of attributes (brand, descriptions) as well as products themselves, the vectors for products spatially lie such that they can be described by using their attributes. Compositions of the vectors would produce a new vector that would resemble to a very relevant product. After these arithmetic operations, we seek to retrieve the top product vectors.

$$\Phi(Attribute1) + \Phi(Attribute2) = \Phi(Product)$$

$$\Phi(Attribute1) + \Phi(Attribute2) - \Phi(Attribute3) = \Phi(Product)$$

miller + lite	**heinz + ketchup**	**kids + shirt**	**phone + charger**
Miller Lite 18/12 C	Heinz Ketchup 44OZ Bonus Pack	Girls Faded Glory	Wall Charger With Cable White
Miller Lite 12/12 C	Heinz Ketchup 38OZ	Hilo Bf Scoop JRS Tee	Charger Combo White
Miller Lite 18/12 B	Heinz Ketchup 14OZ	365Kids Girls Graphic Tee	Wall Charger Combo White

Table 1: Vector algebra with product vectors

miller + lite -12/12	**phone + charger - white**
Miller Lite 18/12 C	Wall Charger
Miller Lite 18/12 B	Color Dual USB Wall Charger
Miller Lite 30/12 C	Universal Wall Charger

Table 2: Vector algebra with product vectors

Table 2 illustrates the behavior of compositionality of three vectors. Just as addition of two vectors combines contexts of the items, subtraction removes the context of the item. Vector arithmetic confirms the ability of the product vectors to aid in faceted product search.

2.3 Substitutes Recommendation

Once representations of all the products are obtained we retrieve the most similar products in the newer product space. It is observed, that most similar items from a particular category also have very low co-occurrence scores. This directly indicates that these products are potential substitutes as can bee seen from Table 3. This resolution is mainly achieved upon the assumption that substitutes are rarely bought together.

Product 1	**Product 2**	**Similarity Score**	**Co-occurrence Score**
Pepsi 2 Liter	Mountain Dew 2 Liter	0.9276	0.0
Pepsi 2 Liter	Dr Pepper 2 Liter	0.8690	0.0
Pepsi 2 Liter	Dr Pepper Diet 2 Liter	0.8195	0.0
Pepsi 2 Liter	Diet Mountain Dew 2 Liter	0.7708	0.0
Pepsi 2 Liter	7UP 2 liter	0.7741	0.0
Pepsi 2 Liter	Coca-Cola 2 liter	0.7316	0.0

Table 3: Potential substitutes, lying closer in the product space

Defining the boundary of the competitive relationships of product pairs is crucial because ill-defined substitutes can negatively impact the sales. Right identification of substitutes can be key in understanding demand transference in case of product out-of-stock and give immense insights for effective pricing decisions. Our results are validated by category experts but we are yet to find an automated way of verifying the competitive relationship between products.

3 Potential Extensions

- Currently the word vectors are obtained from the co-occurrence scores which are solely derived from transaction logs. The current vector doesn't explicitly contain any content-based information of the products (e.g. - brand, description etc.). How a product vector can be created with explicit incorporation of product meta-data along with co-occurrence information by directly modifying the scores will be worth trying.

  To write it explicitly, the co-occurrence score can be augmented with attribute similarity score, user similarity score, price similarity to incorporate these information in the product vectors.

- Usability of these product vectors in product recommendation (complements, 'people also bought') is to be experimented. These product vectors can be used further in sequence learning (alike NLP tasks) to churn out common patterns in shopping. Understanding the temporal dynamics in purchase pattern, shift in product preferences also can be greatly improved upon via using the representation of the products.

- How product vectors improve product search (via vector arithmetic) is to be experimented. The attribute vectors and the product vectors spatially lie in closer space which make the vector arithmetic more interesting. The faceted search can be greatly improved upon by leveraging these vectors in this low-dimensional manifold.

- Dimensional analysis and search for their semantic meaning can unveil the nature of the low dimensional manifold in a more meaningful way. For example, if there is one/more dimension(s) which is(are) responsible for capturing some signal about the product's implicit attribute, a dimensional analysis can be helpful there. If a certain set of items are mostly preferred by a generation e.g. oldies, and if certain dimensions are identified consistently capturing that signal, the similar dimensions can be leveraged well in understanding other products for those this attribute (preference from a specific generation) is not explicit.

- In a vector space, where product vectors and attribute vectors coexist, the sufficiency of considering euclidean distance as measure to capture the proximity is an interesting aspect to explore upon.

- Moreover and finally, generating task-specific product vectors can be an over-arching theme of these experiments. It is already understood that, the product representations for identifying competitive products and for predicting next item to be purchased need separate ways of generating them.

Motivation of this work mostly rests on the recent works (KDD '17) in obtaining vector representations of using metapaths in a heterogeneous network (metapath2vec) or on creating vector representations for node's structural identity (struc2vec). Prolific use of word vectors in various NLP applications evokes the possibility of product vectors being efficient in tasks like search, recommendation, identifying competitive products and many more.