Towards Preprocessing Guidelines for Neural Network Embedding of Customer Behavior in Digital Retail

Douglas Cirqueira
School of Computing
Dublin City University
Dublin, Ireland
douglas.darochacirqueira2
@mail.dcu.ie

Markus Helfert
School of Business
Lero - The Irish Software
Research Center,
Maynooth University
Maynooth, Ireland
markus.helfert@mu.ie

Marija Bezbradica
School of Computing
Dublin City University
Dublin, Ireland
marija.bezbradica@dcu.ie

ABSTRACT

Shopping transactions in digital retailing platforms enable retailers to understand customers' needs for providing personalized experiences. Researchers started modeling transaction data through neural network embedding, which enables unsupervised learning of contextual similarities between attributes in shopping transactions. However, every study brings different approaches for embedding customer's transactions, and clear preprocessing guidelines are missing. This paper reviews the recent literature of neural embedding for customer behavior and brings three main contributions. First, we provide a set of guidelines for preprocessing and modeling consumer transaction data to learn neural network embeddings. Second, it is introduced a multi-task Long Short-Term Memory Network to evaluate the guidelines proposed through the task of purchase behavior prediction. Third, we present a multi-contextual visualization of customer behavior embeddings, and its usefulness for purchase prediction and fraud detection applications. Results achieved illustrate accuracies above 40%, 60%, and 80% for predicting the next days, hours, and products purchased for some customers in a dataset composed of online grocery shopping transactions.

CCS CONCEPTS

• Computing Methodologies • Machine Learning • Machine Learning Approaches • Neural Networks

Keywords

Customer Behavior; Machine Learning; Neural Embedding; Word2Vec; LSTM; Digital Retail

1. INTRODUCTION

Digital retailers are dealing with an evolved omnichannel customer base, which is characterized by their daily generation of data composed of shopping transactions, whether online (E-commerce), mobile, or offline (in-store) [3]. Advancements in digital payment technologies push this reality.

This big data provides retailers with opportunities to analyze and understand customer behavior, enabling applications to engage them with personalized experiences [19] or to protect them against non-legitimate activities on their accounts [13]. Those applications are achieved by the deployment of machine learning models, which learn behavioral patterns out of customer transactions in retail [16].

Recently, some researchers realized the potential of applying sequential machine learning models for analyzing customer transactions, including Recurrent Neural Networks (RNN), and their variants as Long Short-Term Memory Networks (LSTM) [13, 21]. Those models possess memory to consider the temporal order and influence of previous events for providing predictions [10].

Nevertheless, it is challenging to predict consumer behavior transactions. First, such behavior is inherently stochastic, and purchase decisions might be influenced by several factors not possible to control in experiments [7]. Secondly, customer behavior in retail is multi-contextual, meaning it is possible to derive predictive variables from different contexts in shopping transactions, including product details, timestamps, or locations of shopping events [3].

Researchers have tried to overcome such challenges adopting neural network embeddings [18], which is a feature learning method inspired by advancements in the field of natural language processing (NLP).

However, current proposals adopting neural embedding for customer behavior transactions provide a shallow view into their data preprocessing steps and mostly disclose their overall neural network architecture. That is a barrier for benchmarks and advancements in embedding learning for customer behavior, given the lack of standards for their implementation phase. Furthermore, literature usually focuses on predicting a single outcome out of customer transactions, and little is researched on the assessment of embeddings for multi-task customer behavior predictions. Therefore, this paper brings the following contributions:

1. Guidelines on how to preprocess consumer transaction data, and generate multi-context neural network embeddings for learning shopping behavior patterns;

2. A multi-task LSTM for learning neural embeddings for multi-context transactions. We aim to evaluate the embedding approach for predicting the next day of a week, hour and product categories purchased by a customer;

3. A first direction towards visualizing multi-context customer behavior embeddings.

Results illustrate accuracies above 40%, 60%, and 80% for learning customers next days, time of day, and purchase categories, respectively, in a dataset composed of online grocery shopping transactions. The contributions stated can support researchers on the implementation phase of neural embeddings, multi-task predictions, and a novel approach for the visualization of legitimate consumer behavioral patterns. Therefore, the goal is to support the application of neural embeddings to obtain a complete view of customers purchasing patterns, which can impact applications such as to visualize their needs or protect them from fraudulent transactions not matching their shopping behavior.

The rest of this paper is organized as follows: Section 2 brings a literature review containing the proposals mapped for the guidelines proposed; Section 3 presents the main contribution as the neural network embedding guidelines; Section 4 illustrates the experiments and results achieved, followed by Section 5 with conclusions.

2. LITERATURE REVIEW

In order to overcome challenges related to the stochastic and multi-contextual nature of customer behavior, and inspired by NLP advancements, researchers started to adopt neural embeddings for learning shopping transaction patterns [2, 8, 9, 21]. Embeddings are dense vector representations, which map categorical variables to continuous real values or weights, which represent contextual relationships between variables in a dataset, or words in NLP [18]. Those embedding vectors have a specific number of dimensions, equal the number of weights, which is defined by the researcher. Each weight in a dimension may carry information towards a specific context, and for the value it represents. Figure 1 brings a classic example, where words have a stronger relationship in an embedding space due to the gender dimension.
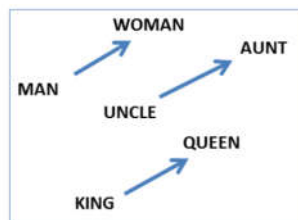

Figure 1. Embeddings related by gender [18].

On the field of customer behavior prediction, researchers try to improve their models by learning those relationships between transaction items in purchase histories.

In [4], the authors propose a recommendation framework to detect sequential purchases based on the embedding of goods, and try to

recommend products based on previous items purchased. In [16], the authors generate embeddings of whole online user sessions, to detect if visitors will perform a purchase. The work of [9] generates embeddings for customer styles in the fashion industry. They apply an LSTM model to detect the next consumers' clothing styles to recommend them. In [24], the authors try to detect if a user's online shopping session will end up with a purchase or cart abandonment. For that, they generate embeddings of online sessions with a mixture of RNNs. [1] proposes a deep learning approach to purchasing histories to detect if customers will repeat orders for products previously bought via deals. Finally, [20] adopts neural embeddings to model the recency, frequency, and monetary variables from customer behavior (RFM). That is the closest proposal to this paper, as they adopt a multi-task neural network architecture, but for predicting the RFM variables. Those proposals have, in common, the embedding strategy identified as Single Transaction Embedding, for which Subsection 3.3 gives further details.

Furthermore, [21] is another example of purchase prediction within an online shopping session. The authors adopted an LSTM to classify users as clickers or buyers. [26] presents an RNN approach to predict and recommend the next locations for customers, based on their previous points of interest and check-ins. The authors on those proposals adopted the second approach identified for embedding customer behavior, which is Multiple Input Embedding (see Subsection 3.3).

The work of [11] illustrates a Gated Recurrent Unit (GRU) based recommender system focused on predicting subcategories of products for next customer purchases, such as accessories for cars. [25] presents the application of neural embeddings on product titles, and successfully employ those for the tasks of product purchase prediction and segmentation of products and customers. In the financing field, the work of [5] uses learned embeddings to detect similarities between banking customers based on their transactional data. They highlight the benefit of adopting embedding for clustering applications. Those proposals have, in common, the application of the third embedding approach identified, which is Pre-Trained Embeddings (see Subsection 3.3).

In summary, there is a diversity of applications for neural embeddings for customer behavior. The authors mostly focus on predicting the next products, locations, orders, or results of an online shopping session. We contribute to the literature by providing guidelines for the preprocessing stage of neural embedding learning, and assess them through a multi-task LSTM, rarely seem in the literature, for predicting the next day of a week, hour and product categories purchased by a customer.

3. NEURAL NETWORK EMBEDDINGS FOR CUSTOMER BEHAVIOR

Through surveying the literature described in the previous section, it was possible to derive the guidelines presented in Figure 2. The critical aspects analyzed in each surveyed paper were: raw data features, preprocessing, data imputation, and processing steps to generate embeddings. Therefore, we derive four steps for preprocessing and learning embeddings for consumer behavior. Those guidelines were tested in the experiments section, which illustrates the performance of the three embedding strategies identified in the literature for purchase prediction, illustrated under the imputation step.
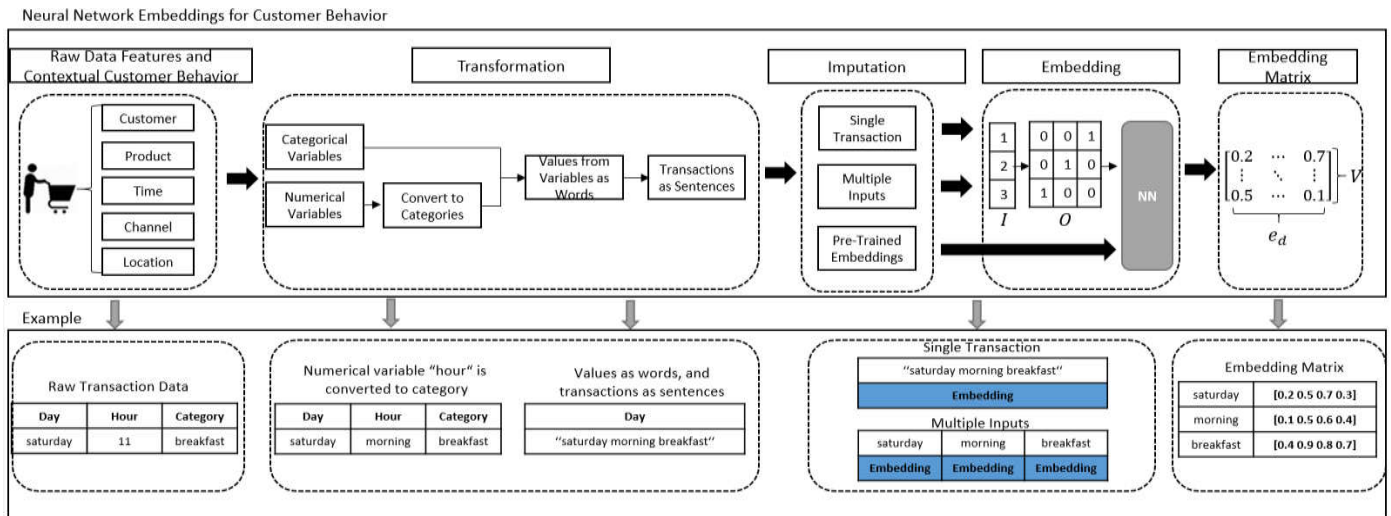
Figure 2. Guidelines for learning neural network embedding of customer behavior.

Table 1. Experiment results for the best hyperparameters of the LSTM model.

Prediction Variable	Day of Week	Hour	Purchase Category
Variable Entropy (How difficult is it to predict this variable)	2.23	1.06	2.85
Accuracy			
Scenario 1: Single Transaction Embedding without week period	**44%**	56%	80%
Scenario 2: Single Transaction Embedding with week period	38%	53%	**82%**
Scenario 3: Single Transaction Embedding with week period and days passed	37%	50%	81%
Scenario 4: Multiple Inputs Embedding without week period	28%	57%	81%
Scenario 5: Multiple Inputs Embedding with week period	30%	54%	81%
Scenario 6: Multiple Inputs Embedding with week period and days passed	29%	50%	81%
Scenario 7: Pre-trained Embeddings (word2vec was applied [18])	32%	**62%**	**82%**

3.1 Raw Data Features

Raw features in consumer transactions are known for being multi-contextual or dimensional [3]. Those dimensions are *Customer, Product, Time, Channel*, and *Location*. Each dimension provides information on consumer shopping behavior and enables the derivation of predictive features from different perspectives.

The *Customer* dimension relates to demographics attributes, such as age or gender. In an online shopping session, such attributes can relate to clicks performed by a customer. The *Product* dimension relates to the raw characteristics of products bought, including category or prices. The *Time* dimension supports the prediction of when events can happen based on previous transaction timestamps and seasonal patterns. The *Channel* dimension enables the assessment of what factors influence customer purchase likelihood when adopting different channels, such as mobile and online.

Finally, the *Location* dimension helps in identifying patterns according to the spatial placement of consumers, such as underlying relationships between their purchases and the influence of their zip code, city, or state location level.

Each of those dimensions might provide categorical or numerical variables, which are handled by the transformation step.

3.2 Transformation

As previously mentioned, to generate neural embeddings of consumer behavior, researchers got inspired by advancements in the field of NLP, which deals with textual data [18]. In the field of consumer behavior, the idea is to visualize transaction attributes as words, and a whole transaction as a sentence.

When dealing with categorical variables, this process is straightforward. However, if dealing with numerical features, it is

necessary to derive categories out of the continuous values that a numerical feature can assume. For instance, some researchers split time stamps into bins representing the four periods of a day, like "morning", "afternoon", "evening" and "night" [21]. Amounts can be split according to their magnitudes, such as "high", "medium", or "low" amount [5]. The choice for deriving those categories is up to the researcher, and ideally, it should try to reflect the contextual meaning that continuous values may carry for an application domain. For example, if a variable represents the age of a customer, it is essential to consider what the company providing the dataset considers as "young" or "adult" before splitting the values according to their customer's profile.

3.3 Imputation

The next step is to provide the transformed data as input to a neural network embedding layer. From the literature identified, it was possible to derive three main strategies for this imputation stage. The first, Single Transaction Embedding, works by providing the predictive customer features as a single transaction to the embedding layer. In this case, all the feature values would have the same embedding dimension [4, 16, 9, 24, 1, 20]. The second strategy, identified as Multiple Input Embedding, works by providing each predictive variable individually to the embedding layer. In this case, the researcher has the freedom to define different embedding dimensions for every feature [21, 26].

Finally, some researchers prefer to work with the Pre-Trained Embeddings strategy [5, 11, 25]. In this case, the input for the model will already be embeddings. In practice, it is possible to load the embedding weights into an embedding layer of a predictive model and reuse them. That is how transfer learning is made possible: generating neural embeddings for a specific task and applying them to a different application [27].

3.4 Embedding

The next step is to learn and generate neural embeddings. It starts with the transformation of categorical feature values to integer values. Each transaction feature value will have a unique integer identifier (vector I in Figure 2). Further, those integer variables are converted to a one-hot encoding representation (vector O in Figure 2). All the cells in this vector will have value zero, except for the one representing the current value a variable is assuming, which is set up to one.

This one-hot encoding vector will be the input to a neural network model, which will learn an embedding matrix. Usually, this matrix is initialized with random weights. Some researchers try to set up the maximum and minimum range those weights can assume [21].

To learn those weights in the NLP field, researchers set up context words, a target word or context sentences, and try to learn the next word, context, or sentence to appear near those inputs, respectively [6, 9]. Words that often appear in the same context will have similar embeddings. In the case of customer shopping transactions, the context or target words is represented by the next values of transaction features. For example, given the attribute values "saturday", "morning" and "home_neighborhood", the probability of a purchase category being "breakfast" can be learned. Therefore, customer transaction attributes that happen to be often in the same context, will have similar embeddings.

When adopting the strategies of Single Transaction and Multiple Input Embedding, the embedding matrix is learned during the training time of a neural network model. In the literature investigated, those neural models are generally RNNs, LSTMs, or GRUs. Then, the embeddings matrix can be optimized through a backpropagation algorithm with a loss function, which guides the model in optimizing the matrix's weights to minimize the error between inputs and outputs, such as:

$$\mathcal{L}(y,x) = \sum_{t=1}^{T} \mathcal{L}(y_t, x_t) \qquad (1)$$

where x_t is the input, y_t output, and T the total of data instances.

The loss function depends on the type of output the neural network model provides. For instance, it should be a binary cross-entropy [14] for binary classification. That is the case when predicting if an online user session will finish with a purchase or not, or if a customer will repurchase a good. On the other hand, the loss function should be a categorical cross-entropy [14] for multiclass tasks, which is the case to detect which product category a user will buy next.

Finally, the learned embedding matrix $M_{V x e_d}$ will have dimension V equal to the total of values assumed by categorical variables (vocabulary), and dimension e_d equal to the embedding output width or dimension defined by the researcher. The embeddings learned in the previous step can be used through different strategies for understanding customer behavior. For example, they can directly serve as inputs to a further classifier, or they can be the inputs to the same neural network model used to learn them, for providing particular outputs.

4. EXPERIMENTS AND RESULTS

The guidelines in Figure 2 are tested, with the main difference from each experiment being the embedding strategies of Single Input, Multiple Input, and Pre-Trained Embeddings. The dataset adopted is the Instacart Online Grocery Shopping from 2017 [23], with 3,421,083 million orders of customers buying groceries through an American online platform. This dataset has 70% of customers rebuying products at least once.

It is selected for experiments the transaction history of the top 10 customers with the highest number of purchases, which gives a total of 30.659 unique orders. The reason is the focus of these experiments on primarily evaluating the embedding strategies identified. The transaction attributes adopted as predictive features are: day of a week ("sunday" to "monday"), hour of the day (split into 4 bins representing "morning", "afternoon", "night" and "dawn time"), and purchase categories (total of 21 categories). Two additional features are tested to enrich the contextual information around the customer behavior and to assess improvement on performance. Those are a feature for the day period on a week ("weekday" or "weekend"), and the number of days passed since the last customer transaction. The last scenario adopts the Pre-Trained Embedding strategy with all the features cited. That results in 7 experimental scenarios. Those predictive features are preprocessed following the guidelines provided in Section 3, as the example in Figure 2 shows. Therefore, the contextual dimensions adopted in these experiments, as illustrated in Subsection 3.1, are *Product* and *Time*.

The neural network model adopted to learn the embeddings is an LSTM, implemented via the Python Keras[1]library. The guidelines given are generalizable so that a researcher can change this model to a GRU or a vanilla RNN without harm. The reason for this choice is the fact that LSTMs have a mechanism to overcome the vanishing gradient problem [12], and to optimize the learning process of long and short-term patterns in sequences. The LSTM is trained with Adam optimizer [15]. A dropout layer with a rate of 20% is employed to avoid overfitting [22]. From every customer, 90% of their orders are used for training, and 10% for testing. After initial experiments, the best settings for the LSTM hyperparameters are 128 hidden units, and an embedding dimension of 20 for Single Transaction Embedding. For Pre-trained Embeddings, the dimension is also 20, with a context window of 5. The word2vec [18] model is applied for this scenario, implemented by the Python Gensim[2]library. For the Multiple Inputs case, the best dimensions are 6 for days of a week, 3 for hour of the day, and 10 for categories. The number of epochs is set to 100.

The LSTM architecture adopted is based on a multi-task classification problem. The aim is to learn and predict the next day of the week, the hour of the day, and purchase category in the customer transaction behavior. Therefore, the target outputs for the model are composed of three variables encoded as one-hot encoding vectors. In the case of purchase categories, a 21-hot encoding vector is used, as multiple categories can be present in an order. The activation functions adopted are categorical cross-entropy for day and hour, and binary cross-entropy for categories.

It was computed the entropy of each target variable from all the orders used for experiments, to detect which would be the most difficult one to predict according to the diversity of choices in the customer transactions. Then, the metric of accuracy is adopted to assess the performance of predictions from the LSTM. Table 1 illustrates those results, from which the purchase category is the most difficult to predict, as expected.

Results in Table 1 illustrate that there are no best settings for predicting all the variables. However, scenario 1 is the best to predict the next day a customer will purchase, with 44% accuracy. For predicting the hour of a purchase and its categories, the best scenario is number 7 with pre-trained embeddings. Settings in scenario 2 perform well for the case of category prediction. Therefore, although some proposals in the literature have achieved higher accuracy than these results, they focus only on one of those variables. Here, it was possible to check the performance when applying a multi-task LSTM for multiple outputs on consumer behavior.

The final experiment focuses on analyzing the usage of pre-trained embeddings, which is the third strategy identified in the literature. In this case, the embeddings learned are given as input and output to an LSTM model. Figure 3 illustrates a t-SNE plot [17] of the embeddings learned for each feature value during training time, and the predicted embeddings for the transaction behavior of the 6th customer in the dataset.
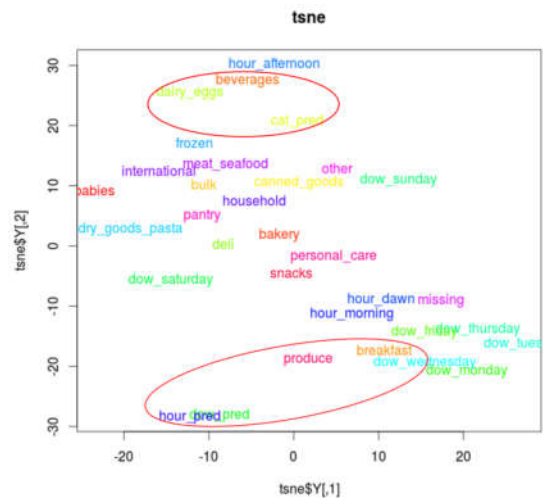

Figure 3. Embeddings as prediction output.

From this chart, it is possible to visualize a novel possibility to analyze customer behavior. It was possible to extract two interesting areas from this chart. The red ellipse on the top indicates the embedding of the category predicted. It is closer to beverages and dairy eggs, which are two real categories purchased by customer 6th on this transaction. The red ellipse on the bottom shows the prediction embeddings for hour of the day and day of week closer to the categories produce and breakfast, which are two other categories acquired by the customer at this transaction.

Therefore, it is possible to visualize the predicted embeddings and their behavioral orientation towards a specific day of a week, hour, or product categories. That opens possibilities to investigate customer transaction patterns, and even protect them from fraudulent transactions, which could show up as outliers in the embedding space.

5. CONCLUSIONS

This paper provides data preprocessing guidelines towards working with neural network embeddings and multi-task predictions to handle shopping transactions and visualize consuming patterns. Embeddings can be robust for customer purchase prediction, as they reveal relationships between multiple dimensions in customer behavior. From the literature review, plus experiments performed, the main findings are:

- There is not a consensus for what is the ideal embedding dimension. Some researchers adopt heuristics, such as it should be at least m-1, where m is the total of values assumed by categorical variables. In the end, it is better to define that through empirical evidence and experiments;

- The addition of more context to predictive embedding features on the same dimension, or numerical variables combined with embeddings, did not improve accuracy as it was expected. That is the case of the week period and days passed variables;

- Multi-task settings are challenging for predictive models. However, Single Transaction and Pre-Trained Embeddings are promising for multi-task prediction with LSTMs for

---

[1] https://keras.io/

[2] https://radimrehurek.com/gensim/index.html

customer behavior. We consider a further investigation of the combination of both approaches;

- Literature mentions the need for a massive dataset for learning embeddings. In fact, it seems 30.000 transactions were not sufficient to detect significant performance differences between the three embedding strategies identified. However, it was still possible to obtain reasonable accuracy, over 60% and 80%, with a smaller dataset for predicting the hour and purchase categories, plus visualization of customer behavior in the embedding space. That reinforces the potential of this approach for consumer behavior learning;

- The usage of embeddings increases the range of possibilities for outputs from predictive models for customer behavior, beyond binary outcomes and probabilities. Their visualization enables the detection of relationships from products to products, products to times, and times to times. For instance, that can enable the multi-context understanding of shopping patterns such as products bought together and in specific times.

As future work, it will be performed an investigation of multi-context embeddings in learning consumer behavior, considering the impact of additional dimensions, such as location and customer attributes. It is planned to increase the number of customers and dataset size for evaluating the embedding approaches. Although embedding weights can reveal hidden relationships between variables, they are not interpretable. That is a potential research direction in this area for consumer behavior. Finally, it would be interesting to evaluate these guidelines for transaction fraud detection in retail.

ACKNOWLEDGMENTS

REFERENCES

[1] Anand, Gaurangi, et al. "Deep temporal features to predict repeat buyers." NIPS 2015 Workshop: Machine Learning for eCommerce. 2015.

[2] Barkan, Oren, and Noam Koenigstein. "Item2vec: neural item embedding for collaborative filtering." 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2016.

[3] Bradlow, Eric T., et al. "The role of big data and predictive analytics in retailing." Journal of Retailing 93.1 (2017): 79-95.

[4] Chen, Xu, et al. "Sequential recommendation with user memory networks." Proceedings of the eleventh ACM international conference on web search and data mining. ACM, 2018.

[5] Dayioglugil, Ali Batuhan, and Yusuf Sinan Akgul. "Continuous Embedding Spaces for Bank Transaction Data." International Symposium on Methodologies for Intelligent Systems. Springer, Cham, 2017.

[6] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[7] Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. Journal of Business Research, 69(2), 897-904.

[8] Guo, Cheng, and Felix Berkhahn. "Entity embeddings of categorical variables." arXiv preprint arXiv:1604.06737 (2016).

[9] Heinz, Sebastian, Christian Bracher, and Roland Vollgraf. "An LSTM-based dynamic customer model for fashion recommendation." arXiv preprint arXiv:1708.07347 (2017).

[10] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[11] Huang, Jin, et al. "Taxonomy-aware multi-hop reasoning networks for sequential recommendation." Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. ACM, 2019.

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning . MIT Press. http://www.deeplearningbook.org.

[13] Jurgovsky, Johannes, et al. "Sequence classification for credit-card fraud detection." Expert Systems with Applications 100 (2018): 234-245.

[14] Keras.io. (2019). Losses - Keras Documentation. [online] Available at: https://keras.io/losses/ [Accessed 14 Jul. 2019]

[15] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[16] Lang, Tobias, and Matthias Rettenmeier. "Understanding consumer behavior with recurrent neural networks." Workshop on Machine Learning Methods for Recommender Systems. 2017.

[17] Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning research 9.Nov (2008): 2579-2605.

[18] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.

[19] Rodríguez-Torrico, P., Cabezudo, R. S. J., & San-Martín, S. (2017). Tell me what they are like and I will tell you where they buy. An analysis of omnichannel consumer behavior. Computers in Human Behavior, 68, 465-471.

[20] Salehinejad, Hojjat, and Shahryar Rahnamayan. "Customer shopping pattern prediction: A recurrent neural network approach." 2016 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2016.

[21] Sheil, Humphrey, Omer Rana, and Ronan Reilly. "Predicting purchasing intent: automatic feature learning using recurrent neural networks." arXiv preprint arXiv:1807.08207 (2018).

[22] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The journal of machine learning research 15.1 (2014): 1929-1958.

[23] The Instacart Online Grocery Shopping Dataset 2017", Accessed from https://www.instacart.com/datasets/grocery-shopping-2017 [Accessed 07 Jun. 2019]

[24] Toth, Arthur, et al. "Predicting Shopping Behavior with Mixture of RNNs." eCOM@ SIGIR. 2017.

[25] Woltmann, Lucas, Maik Thiele, and Wolfgang Lehner. "Modeling Customers and Products with Word Embeddings from Receipt Data." Proceedings of the 22nd International Database Engineering & Applications Symposium. ACM, 2018.

[26] Xia, Bin, et al. "Attention-based recurrent neural network for location recommendation." 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE, 2017.

[27] Zhang, Li, Tao Xiang, and Shaogang Gong. "Learning a deep embedding model for zero-shot learning." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.