An intelligent shopping list based on the application of partitioning and machine learning algorithms

Nadia Tahiri[1], Bogdan Mazoure[2] and Vladimir Makarenkov[1]

[1]Université du Québec à Montréal and [2]McGill University

Objectives

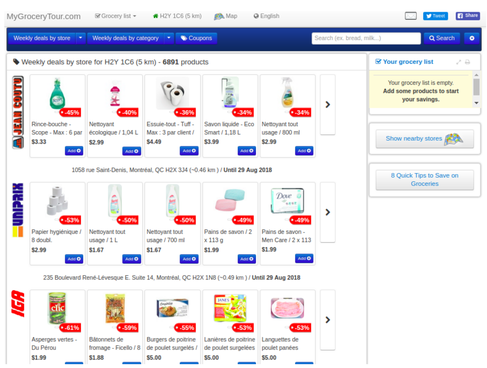
- Predict which groceries the consumer will want to buy again or will try to buy for the first time, and in which store(s) in a given area he will shop;
- Create a smart shopping list by offering the consumer a personalized weekly shopping list based on their shopping history and current promotions;
- Depending on the user u and the user's purchase history ($\text{order}_{t-h:t}$, $h > 0$), predict the probability that a product i is included in the next order_{t+1} of u.

Introduction

A shopping list is an integral part of the shopping experience for many consumers. A typical grocery retailer offers consumers thousands of promotions every week to attract more consumers and thus improve their sales [1].

Dataset

Figure 1: Website *CircuitPromo.ca* with the postal code H2Y 1C6 in Montreal



Methods

Representation	Description	Type
Products	Model $\text{P}(\text{product}_i \in \text{order}_{t+1})$ with $\text{orders}_{t-h,t}$, $h > 0$.	LSTM (300 neurons)
Categories	Predicts $\text{P}(\exists i : \text{product}_{i,t+1} \in \text{category}_r)$.	LSTM (300 neurons)
Size	Predict the size of the order_{t+1}.	LSTM (300 neurons)
Users Products	Decomposed $V_{(u\times p)} = W_{(u\times d)}H^T_{(p\times d)}$	Dense (50 neurons)

Table 1: Top-level models used.

Model	Embedding	Dimensions
LSTM Products	Products	$49,684 \times 300$
LSTM Products	Categories	24×50
LSTM Products	Users	$1,374 \times 300$
NNMF	Users	$1,374 \times 25$
NNMF	Products	$49,684 \times 25$

Table 2: Dimensions of the representations learned by different models.

Products/Baskets

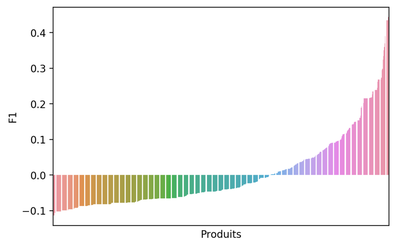


Figure 2: Distribution of F_1 measures relative to products, around average.

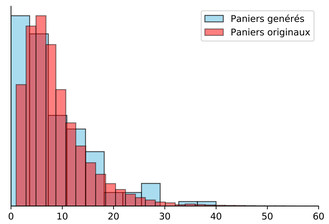


Figure 3: Distribution of the size of the predicted and original baskets.

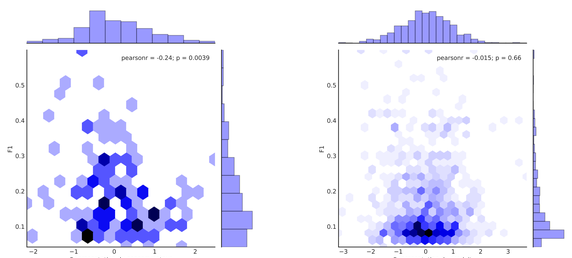
Validation step

The final basket is chosen according to the final probabilities of reorganization, choosing the subset of products with the maximum expected F_1 score [2].

$$\max_{\mathcal{P}} \mathbb{E}_{p' \in \mathcal{P}}[F_1(\mathcal{P})] = \max_{\mathcal{P}} \mathbb{E}_{p' \in \mathcal{P}}\left[\frac{2 \sum_{i \in \mathcal{P}} \text{TP}(i)}{\sum_{i \in \mathcal{P}} (2\text{VP}(i) + \text{FN}(i) + \text{FP}(i))}\right],$$

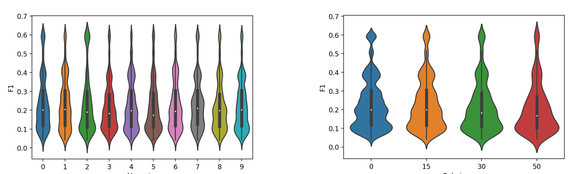
where $\text{TP}= \mathbb{I}[\lceil p(i) \rceil = 1]\mathbb{I}[R_i = 1]$,
$\text{FN}= \mathbb{I}[\lfloor p(i) \rceil = 0]\mathbb{I}[R_i = 1]$,
$\text{FP}= \mathbb{I}[\lceil p(i) \rceil = 1]\mathbb{I}[R_i = 0]$ and
$R_i = 1$ if the product i was bought in the cart $p' \in \mathcal{P}$, otherwise 0.
We was used $\mathbb{E}_X[F_1(Y)] = \sum_{x \in X} F_1(Y = y|x)P(X = x)$

Consumers



(a) (b)

Figure 4: Distribution of F_1 measures relative to consumers (a) and products (b).



(a) (a) (b) (b)

Figure 5: Distribution of F_1 measures against stores (a) and rebates (b).

Conclusion

- We modeled the habits of consumers on *CircuitPromo.ca* by using deep neural networks.
- We used two types of neural networks during learning: recurrent neural networks (RNN) and Feedforward neural networks.
- The value of the F_1 statistic that represents the quality of our model is 0.22. The constant influx of new data on *CircuitPromo* will improve the model over time.
- The originality of our approach, compared to existing algorithms, is that in addition to the purchase history we also consider promotions, possible purchases in different stores and the distance between these stores and the consumer's home.

References

[1] Arry Tanusondjaja, Magda Nenycz-Thiel, and Rachel Kennedy. Understanding shopper transaction data: how to identify cross-category purchasing patterns using the duplication coefficient. *International Journal of Market Research*, 58(3):401–419, 2016.

[2] Zachary C Lipton, Charles Elkan, and Balakrishnan Naryanaswamy. Optimal thresholding of classifiers to maximize f1 measure. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 225–239. Springer, 2014.

Contact Information

- Web: CircuitPromo.ca
- Email: nadia.tahiri@gmail.com
- Phone: +1 (514) 629 1474

UQÀM

Fonds de recherche sur la nature et les technologies Québec