Improving Real-Time Grocery Stock Accuracy using Crowd Sourcing and Machine Learning

Derek Yang
University of Illinois
Urbana-Champaign
mcy3@illinois.edu

Yoonhwa Jung
University of Illinois
Urbana-Champaign
yoonhwa2@illinois.edu

Haozhe Wu
University of Illinois
Urbana-Champaign
haozhew3@illinois.edu

Yonnie Kang
University of Illinois
Urbana-Champaign
sk129@illinois.edu

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract

With the popularization of mobile apps for every grocery store, grocery shopping experience has been largely boosted because of the provided pick-up or delivery functions. However, the inconsistency between the availability of products between the physical store and the mobile app always exists, especially during shopping rush hours such as days before college semester starts and days before federal holidays. This paper proposes a crowd sourcing based shopping app which allows users to manually update the availability of products, and shopping recommendations will be provided based on the user labels.

Author Keywords

Human-computer Interaction; Crowd Sourcing; Machine Learning; classification model.

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI);** *Interaction paradigms;* User studies;

Introduction

Almost every large grocery store in the U.S. has an app. Users can use it for locating certain products in the store, prepaid pick-up, and even delivery. Unfortunately, the availability of products on apps are sometimes not being updated in time, especially during shopping rush hours. This might result in frustrating shopping experiences when users drove to the grocery store and found out the products were out of stock despite being informed by the app they were in stock. Similar scenarios also happen in online orders. Users made the order once being informed by the app that the products were available, but the products were either refunded or replaced by other items later. This has caused unnecessary waste of time.

Multiple reasons may lead to such inconsistency. It is difficult to sync the precise stock data of products on the shelf with the app. Most of grocery stores keep track of inventories by using stock management software, where products need to be manually scanned into the system when restocking happens. When the number of products gets huge, accuracy will not be guaranteed in this process. In addition, when products are purchased, the stock updates based on the transactions. Customers may later return the products, or products are placed into the wrong shelf by customers and the store loses track of them, or the products were simply placed into karts and have not been checked out yet. All these complicated factors can result in inaccurate stock data on the apps. Similarly, there were six composite reasons for inventory discrepancies in the warehouse. Similarly, according to a preliminary finding of warehouse stock discrepancy root causes [8], it has been observed that environmental and manpower challenges can contribute to inaccurate stock data. The study provides additional evidence supporting the difficulties in synchronizing stock data, manual scanning processes, and various factors leading to inconsistencies in inventory records.

To improve the stock accuracy for grocery shopping, this paper proposes a design of a crowd-sourcing based grocery app. Such app allows the customers to manually label the product availability whenever they go shopping. The back-end will make smart recommendations to other users based on customer's labels, time stamps, and user credits.

Due to lack of time and labor, the team decided to collect a small portion of data using crowd-sourcing and generate the rest of data based on a series of ground truth for model training. To simulate the crowd-sourcing data collection, we four people and two other friends took notes about the product availability whenever we went to the grocery store. We mainly focused on false positive (actually out of stock but marked as in stock) and true negative cases (both the shelf and the app are out of stock). It was because it is believed that these two cases affect the user shopping experiences the most. The true positive and false negative cases, though also play important roles, are much less affecting. We in total collected more than 200 data points in 8 different grocery stores in Champaign, Illinois. Among the 131 out of stock products, 78.6% of them were actually marked as in stock on apps. It means that the false positive rate is high in real world grocery apps. With crowd-sourcing, users of the app will be notified by the labeling of the customers who discovered the unavailability of the products.

In order to train the machine learning model for smart recommendation, the team decided to generate data to compensate for the lack of labeled data. We generated 378,000 data points based on a series of ground truths.

Background

The topic of out-of-stock events has been widely explored for proper retail operations and user satisfaction. With the increasing use of real-time supermarket shopping services, customers are able to easily track the availability of items and make a order through a supermarket retail mobile apps. Nevertheless, the accuracy of real-time stock is not always accurate reflecting real-time information. This directly affects to customer satisfaction. In particular, managing real-time supermarket stock data is challenging since it is difficult to synchronous the store aisles data with app database.

The real-world practice on the detection of out-of-stock events is a manual check system by employees (physical audits) [5]. To alleviate this challenge, several research implemented an Radio-Frequency Identification technology (RFID) system for monitoring on-shelf availability [3, 7]. Chuang et al. [4] used transaction data shared by the retailer to detect out-of-stock events and correct them with physical audits. These research enables to monitor the availability of on-shelf products and refill the items. As a more recent approach, Allegra et al. [1] utilized images captured by a camera mounted in shopping carts to detect out-of-stock items.

However, predicting out-of-stock events is important to manage supply chains and improve customer's satisfaction, beyond just detecting out-of-stock items. To do that, Andaur et al. [2] predict out-of-stock using Random Forest and Ensemble (ENS-Stack) of the previous six classification algorithms using historical point-of-sale data with stores' physical audits data. Giaconia and Chamas [6] proposed a deep learning-based method for predicting the residual stock amount of a product based on visual counting system, management database, and climate analysis systems. These methods rely on multiple data sources and do not consider human-interaction in the prediction process. Specifically, they do not consider widely used supermarket app availability that users can look up certain product locations in a store, set prepaid pick-up, and even delivery.

To bridge current problems, we explore ML classification models for increasing the accuracy of real-time stock by collecting data from crowd sources based shopping app which allows users to manually update the availability of products. Building on the prediction models, we aim to design a app that enables this process that can gather all the local grocery store data to make smart recommendation to users, avoid browsing across multiple apps.

Dataset

Proof of concept

Fig. 1 shows examples of our crowed-sourced dataset. In our dataset, we have gathered information including the market name, item category, product name, price, actual availability, availability on the app, data timestamp, and the day of the week. To collect this data, we play as the role of crowd workers and collected information from eight different grocery stores in Urbana and Champaign, IL over a period of two weeks.

An interesting finding from our analysis is about the false positive and true negative cases. We observed that among the 131 items that were out-of-stock, there was a 78.6% false positive rate. This implies that numerous items were actually unavailable, but they were still being shown as available on the supermarket application.

This finding highlights the importance of utilizing crowd-sourced data gathered from users along with an effective way of verifying such data. By involving users in the data collection process, we can effectively reduce false positive cases in real-world scenarios. Additionally, the credit and reputation of the users can act as a deterrent for malicious users, further enhancing the reliability of the collected data.

Generated Data

Due to lack of time and labor, we decided to create dataset based on a series of ground truth (following predefined probability rules) for model training, on top of our collected data mentioned in Section. Crowed-Sourced Data .

id	Market Name	Category	Product	Price	Actual Availability	Availability on APP	Data	Timestamp (	Day of the week	User's Credit
	Urbana Shnucks	Fruits	Blueberry	2.25	Less in stock	In stock	4/15	11:20	Sat	
	Green Target	Fruits	Org Blueberries 6oz	3.19	Out of stock	In stock	4/15	19:14	Sat	
	Green Target	Fruits	Org Strawberries 16oz	5.89	Out of stock	Out of stock	4/15	19:20	Sat	
	Green Target	Vege…	Org Broccoli Florets 12oz	4.99	Out of stock	In stock	4/15	19:22	Sat	
	Green Target	Daily…	GG Milk Slim Fat Free	2.39	Out of stock	In stock	4/15	19:26	Sat	
	Green Target	Daily…	GG Milk Whole Vit D	1.79	Out of stock	Out of stock	4/15	19:28	Sat	
	Green Target	Nutrit…	24.8oz PZROL	8.69	Out of stock	Out of stock	4/15	19:32	Sat	
	Green Target	Dairy	Ben&Jerry Respberry cheese cake	5.29	Out of stock	Out of stock	4/15	19:34	Sat	
	Green Target	Dairy	my/mochi vanilla icecream	5.39	Out of stock	In stock	4/15	19:36	Sat	
	Green Target	Dairy	my/mochi chocolate icecream	5.39	Out of stock	In stock	4/15	19:37	Sat	
	Green Target	Brea…	Argo Corn Starch	2.29	Out of stock	In stock	4/15	19:38	Sat	
	Green Target	Meal	Meal Kit BBQ	16.99	Out of stock	Out of stock	4/15	19:40	Sat	
	Green Target	Dairy	GG Milk Slim Fat Free		Out of stock	Out of stock	4/16	18:00	Sun	
	Green Target	Dairy	GG Milk Whole Vit D		Out of stock	Out of stock	4/16	18:00	Sun	
	Green Target	Fruits	Org Strawberries 16oz		Out of stock	Out of stock	4/16	18:00	Sun	
	Green Target	Fruits	Bannana		Less in stock	Out of stock	4/16	18:00	Sun	
	Urbana Shnucks	Vege…	Spinach	3.89	Out of stock	In stock	4/16	18:30	Sun	
	Urbana Shnucks	Vege…	Radishes	1.89	Out of stock	Many in stock	4/16	18:30	Sun	
	Urbana Shnucks	Vege…	Red Butterhead Lettuce	3.34	Out of stock	In stock	4/16	18:30	Sun	
	Urbana Shnucks	Vege…	Schnucks Whole Brussels Sprouts	3.49	Out of stock	In stock	4/16	18:30	Sun	

Figure 1: Examples of Crowed-Sourced Data

Ground Truth

In our dataset, we have 9 product categories: Dairy, Fruits, Vegetables, Daily necessities, Beverages, Meat, Snack, Personal Care, and Home supply. Each of the categories has a probability of being out of stock for that specific week, 85% for Dairy, 90% for Fruits, 70% for Vegetables, 40% for Daily Necessities, 40% for Beverages, 90% for Meat, 50% for Snack, 30% for Personal Care, and 20% for Home supply.If said category would be out of stock for that week, we have a set of out-of-stock rate which represent the probability that the category would be out of stock at that day, Sunday 4%, Monday 9%, Tuesday 14%, Wednesday 14%, Thursday 18%, Friday 18%, Saturday 23%.Also, we consider that a supermarket restocks every Sunday. Base on the following ground truth, we can then generate our machine training data.

Training DataSet

The model takes in sets of variable, namely, user credit, Category, availability, and timestamp. To generate such dataset, we build the training dataset base on the ground truths we predefined. For a span of 10 weeks, we will generate 100 reports from 100 different user every 4 hours for every category. Every report would be either a false information e.g. not telling the truth, or a true information e.g. telling the truth. The user credit score would we the deterministic factor of whether this user would be telling the truth, we pull this credit score base on a normal distribution $(\mu = 4.0, \sigma = 0.7)$ to simulate a random crowd. The credit score would then be multiplied by 0.2 to translate to a probability of telling the truth. For example, a user with a credit score of 4.0 would have a probability of 0.8, which means this user has a 80% chance of telling the truth. The user would then report the category's information base on whether the user is telling the truth and the actual ground truth of that certain category at the current time stamp. In total, we would have 378,000 data points for our model to be trained on.

ML Models and Results

Decision Tree

Taking into account the nature of our problem, we think that a Decision tree model would be the most appropriate fit. From a logical standpoint, as more time elapses since the restock, the probability of an item being out of stock increases. Moreover, various categories of items have distinct probabilities of going out of stock. For instance, perishable goods such as dairies and meat are more prone to going out of stock sooner compared to home supplies.

At first, we trained the tree without any restrictions which made the model over fit with our dataset because we didn't restrict the maximum depth of the tree. The model reached a depth exceeding 30 during training, and unfortunately, it captured some atypical patterns that are undesirable for our purposes, instead of checking categories and time of week, the models is checking certain number of weeks which means the model learned the ground truth of that week and was making decisions accordingly.

To address this issue, we limited the maximum depth of the tree to 5. The result of said tree as shown in Fig. 2 came out perfectly as we expected, the model is checking the category type and day of week on the higher levels and making decisions accordingly. The model turned out to have a 86% accuracy with a recall of 0.9 and precision of 0.9.

Gaussian Naive Bayes

We also trained the data with Gaussian Naive Bayes model. As a variant of Naive Bayes, Gaussian Naive Bayes assumes that each class follow a Gaussian distribution and supports continuous data with independence assumptions between features. This model achieves 0.89% of accuracy on the out-of-stock prediction.

App Design

The app will offer the choice of grouping the products by categories or markets. Users can search for specific products. Clicking each product will open a list of listing information. Users can choose either "In Stock" or "Out of Stock" option next to the listing information. The app itself will utilize the selections of users to make recommendations of grocery stores to other users for the purchase of the products. There will be a score under the "In Stock" and the "Out of Stock" options, which will be weighted by the number of users who chose such an option and their corresponding credit score.

User Credibility

We suggest making the user credit score a dynamic measurement of a user's credibility. The labeling of a user with higher user credit will be more reliable. A "reasonable" labeling will slightly increase the user credit, whereas a "malicious" operation will result in a penalty on the user's credibility. Behaviors such as marking 100 products as unavailable in a short time frame may be attributed as a "malicious" operation by the system, and thus, certain number of points will be deducted on the user credit. Credibility can also be measured by comparing other users' responses to the same stocks. A higher credit score can be given if a user's response matches many other users' responses.

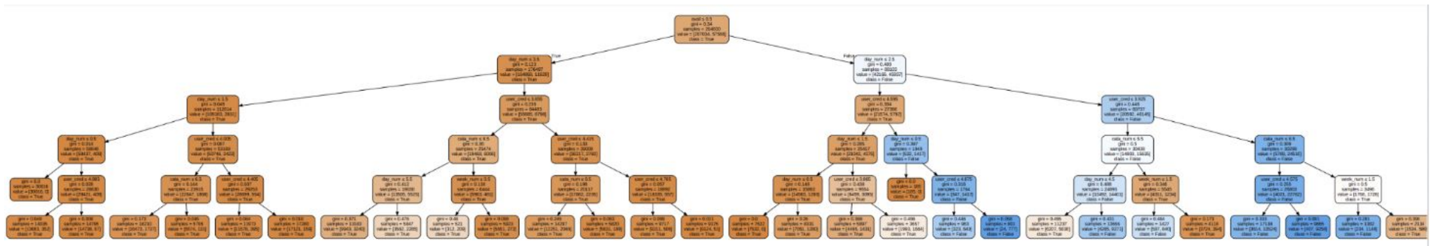

Figure 2: Decision Tree for the out-of-stock prediction

User Incentive

To attract more users to use the app, we can cooperate with grocery stores. Users making a certain number of "reasonable" labeling can be rewarded with grocery store coupons. In addition, leveraging users' reputations can be a way to further enhance user incentives. Implementing a ranking system based on user credits and region can create a sense of competition and recognition among users. Showcasing the top users within the community based on their neighborhood, not only motivates them to maintain their reputation but also can inspire others to actively participate and contribute.Consideration of these user incentives facilitates data collection, which can significantly contribute to the overall accuracy and reliability of real-time inventory data of the app.

Discussion

Our study aimed to address the challenges posed by inaccurate supermarket stock data through the combination of crowd-sourcing and ML model. Also, we suggested the design of the app which can achieve this approach. The analysis of this data revealed a significant discrepancy in the accuracy of current stock information. With a false positive rate of 78.6% observed from just over 200 data points, it shows that there is room for improvement in the reliability of supermarket stock data.

To overcome these challenges, we explored the ML-based approach, specifically the decision tree model and Naive Bayes, using Ground Truth-based generated data. Through this result, we can confidently conclude that ML-based methods work effectively and efficiently on the prediction task even when there is potential malicious player in the system. The model accompanied with crowd sourcing data, could potentially be significantly improved by out method. The model could sense out potential malicious users. This can be a significant contribution, as it demonstrates the robustness and resilience of these models in real-world scenarios.

The design of the app we presented offers valuable insights into the integration of crowd-sourced data into ML-based models. It addresses important considerations such as the collection of actual data and the reliability of such data. Our proposed system's dynamic credit score measurement method and the ability to differentiate between malicious and genuine user responses provide a foundation for realizing this integration.

Limitation and Future Work

At the moment, we have taken on the task of generating both the training dataset and the validation dataset ourselves. However, to ensure a thorough validation process and establish a solid foundation for justifying the effectiveness of our model, it is essential to acquire a dataset that is collected from a diverse crowd. Such a crowd-sourced dataset would offer a broader range of perspectives, experiences, and inputs, allowing us to evaluate the model's performance in a more comprehensive manner. By incorporating a crowd-collected dataset into our validation efforts, we can enhance the reliability and credibility of our findings, reinforcing the robustness of our model's capabilities.

Conclusion

In conclusion, this study delves into the utilization of crowd-sourcing and ML classification model to enhance the accuracy of supermarket stock. Also, We propose a comprehensive design of the app to facilitate the integration of these approaches. Throughout our research, we collected the false positive rate of supermarket stock discrepancies, and developed and tested an ML model, achieving an accuracy of 86%. Even though we used generated data to test it, it still shows our approach can work to solve the problem. Additionally, we presented a design of the app that can seamlessly incorporate ML models.

We hope that this study will contribute to the ongoing efforts to improve the user experience through the combined utilization of ML and crowd-sourcing techniques for enhancing the accuracy of supermarket stock. By adopting this approach, we expect that it can be possible to improve the user's supermarket shopping experience and satisfaction.

REFERENCES

[1] Dario Allegra, Mattia Litrico, Maria Ausilia Napoli Spatafora, Filippo Stanco, and Giovanni Maria Farinella. 2021. Exploiting Egocentric Vision on Shopping Cart for Out-Of-Stock Detection in Retail Environments. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1735–1740.

[2] Juan Manuel Rozas Andaur, Gonzalo A Ruz, and Marcos Goycoolea. 2021. Predicting out-of-stock using machine learning: an application in a retail packaged foods manufacturing company. *Electronics* 10, 22 (2021), 2787.

[3] Eleonora Bottani, Massimo Bertolini, Antonio Rizzi, and Giovanni Romagnoli. 2017. Monitoring on-shelf availability, out-of-stock and product freshness through RFID in the fresh food supply chain. *International Journal of RF Technologies* 8, 1-2 (2017), 33–55.

[4] Howard Hao-Chun Chuang, Rogelio Oliva, and Sheng Liu. 2016. On-shelf availability, retail performance, and external audits: A field experiment. *Production and Operations Management* 25, 5 (2016), 935–951.

[5] Marshall Fisher and Ananth Raman. 2010. *The new science of retailing: how analytics are transforming the supply chain and improving performance*. Harvard Business Review Press.

[6] Concetta Giaconia and Aziz Chamas. 2023. Innovative Out-of-Stock Prediction System Based on Data History Knowledge Deep Learning Processing. *Computation* 11, 3 (2023), 62.

[7] Christian Metzger, Frédéric Thiesse, Stanley Gershwin, and Elgar Fleisch. 2013. The impact of false-negative reads on the performance of RFID-based shelf inventory control policies. *Computers & Operations Research* 40, 7 (2013), 1864–1873.

[8] Sahbudin Sahbudin and Dendi P. Ishak. 2022. A Preliminary Finding of Warehouse Stock Discrepancy Root-Causes: A Case of PT Astra International Tbk Indonesia Warehouse. In *Proceedings of the 4th Asia Pacific Conference on Research in Industrial and Systems Engineering (APCORISE '21)*. Association for Computing Machinery, New York, NY, USA, 583–589. DOI:http://dx.doi.org/10.1145/3468013.3468661