Smart^{Healthy} Coupons: A dynamic coupon model that fulfills user preferences with healthier options

A. Ben Kertman
Harvard School of Public Health
677 Huntington Avenue
benkertman@mail.harvard.edu

ABSTRACT
In this paper we will discuss the increasing costs of the obesity epidemic and how current dynamic coupon systems are likely exacerbating the issue. We propose a solution by sorting dynamically recommended coupons by the healthfulness of each product using an algorithm designed by NuVal, Inc. We explain in detail how this system would operate; how the user experiences these coupons; the design decisions for working with manufacturers, retailers and stores; levels of evaluating the system; possible issues that could arise; what I have learned along the way; and potential next steps.

Author Keywords
Recommender system, big data, obesity, health disparity, dynamic coupons

ACM Classification Keywords
G.3; H.3.3

THE PROBLEMS

Obesity
One of the greatest public health issues that society and especially low-income populations face today is the rapid rise of obesity and obesity related diseases [1]. The list of factors that are causing this epidemic in low-income populations are complex and diverse:

• Unhealthy food costs less, so people on a strict budget who have to maximize quantity must sacrifice nutritional value [2].

• Budgets become so restricted that individuals eat less and then overeat upon receiving the next paycheck. This type of eating behavior can lead to weight gain [3].

• Increased levels of stress due to low-socioeconomic status

factors can lead to unhealthy eating behaviors as well as disruptions in hormonal and metabolic processes, exacerbating weight gain [4].

• Targeted marketing to children and adolescents (ie cartoon characters on boxes of sugar filled cereals) in the supermarket space is likely to lead to greater demand for unhealthy products.

• Low-income individuals often work multiple, extending their total hours outside of the house. This can make it difficult to find time to prepare a meal, leading to the consumption of fast foods and already prepared meals [5].

• A lack of education regarding how to prepare meals leaves few alternatives for obtaining the required sustenance, often leading to unhealthy choices.

The long-term physical health, mental health, and economic effects of childhood obesity have been documented in great detail. Childhood obesity has been shown to be a major factor in morbidity and mortality in later life; increasing "rates of diabetes, coronary heart disease, atherosclerosis, hip fracture and gout [6]. Additionally, obese children are more likely to have a lower level of quality of life (QOL) then children with cancer, an indication of lower "emotional, social, and school functioning"[7].

The Filter Bubble
Another issue, perhaps equally as potent, is a concept that has arisen with technologies that source information for the individual: the Filter Bubble [8]. Before explaining how this issue relates to the obesity epidemic, it might be useful to understand the Filter Bubble by which it applies to a topic more well known; Netflix. When Netflix filters information, they are doing it based upon the individual's preferences as they are in this moment and what they have been in the past. Netflix identifies attributes of a film (director, actors, genre, run-time, etc.), analyzes other films that people with similar tastes have watched, and use both those variables to design an array of films or television shows that the individual is likely to enjoy. The Filter Bubble problem arises because Netflix filters out all other content that is unrelated to the users' preferences, thereby reinforcing the individual's current movie/television watching habits. As this phenomenon continues, the least popular styles of shows and movies become less likely to be

selected and eventually fall out of the option tray, while the user becomes further isolated from alternative options outsider their own preferences. This also occurs with Google searches and on websites such as Reddit, where the least viewed (most down-voted) material falls from view of other people, further making it difficult for people to select it later [8]. Essentially, minority ideas/preferences/options are less likely to receive exposure and fade to obscurity while predispositions of preferences are reinforced; limiting diversity of thought and subsequently behavior.

The Filter Bubble Problem as it Relates to Public Health
The Filter Bubble also applies to food options, specifically with the case of dynamic coupon models. Dynamic coupons for grocery shopping are coupons that are recommended to specific individuals based upon their shopping preferences. The preferences are traditionally identified through the membership card of the store. These coupons are recommended much in the same way that Netflix recommends movies to watch.

The problem is that the more a coupon is recommended based upon current eating habits, the more those eating habits are reinforced. Since low-income populations are already more likely to eat less healthy food for reasons mentioned above, dynamic coupons are likely to exacerbate this effect. In addition, low-income populations utilize coupons in greater numbers than middle or high-income individuals, which is likely contributing to the increase in health disparities [9]. In other words, the populations that need greater assistance to eat healthier become further engrained in their own unhealthy habits while wealthier individuals are less likely to have their current eating behaviors reinforced.

Besides the impending social, economic and health costs due to obesity, expanding health disparities between social classes is at the core of many of today's societal problems. Health disparities and premature deaths alone cost the U.S. $1.24 trillion between 2003 and 2006 [10].

SOLUTION
The Filter Bubble is an example of how recommendation systems, no matter what, cannot be devoid of influence and, in fact, are biased towards very specific behaviors and outcomes. These recommendation systems necessitate other variables that can provide an option towards healthier behaviors rather than only towards reinforcement of current behaviors.

I propose a dynamic coupon model called Smart[Healthy] Coupons that is adjusted towards healthier food items while keeping within the individual's preferences. Each customer will receive individually tailored coupon offers printed upon their receipt every time they check out of a grocery store. In order to explain how this works I will first outline the design of the recommender system and then how it can be adjusted to provide healthier options for the individual.

The Recommender System
There are two primary filtering techniques that are used to create recommendations for customers; content filtering and collaborative filtering. Content filtering evaluates the attributes of a customer's profile (i.e. age, gender, ethnicity, income, membership ID #, etc.) and compares it with the attributes of all the products that person has purchased in the past (brand, price, location in the store, type of food, etc.) [11]. By identifying an individual's preference for those attributes, the system is able to identify all other products that the user does not purchase but also fulfills as many of the attribute preferences as possible. The items that overlap the most are the ones selected as a recommendation. Below you can see the algorithm that is used to determine the utility (u) of each product (s) for each customer's preference profile (c). This technique is called the Informational Retrieval Method and it is executed through term frequency – inverse document frequency (tf-idf) [11]:

u(c,s) = score(ContentBasedProfile (c), Content (s))

$$u(c, s) = \cos(\vec{w}_c, \vec{w}_s) = \frac{\vec{w}_c \cdot \vec{w}_s}{||\vec{w}_c||_2 \times ||\vec{w}_s||_2}$$

$$= \frac{\sum_{i=1}^{K} w_{i,c} w_{i,s}}{\sqrt{\sum_{i=1}^{K} w_{i,c}^2} \sqrt{\sum_{i=1}^{K} w_{i,s}^2}},$$ [11]

The collaborative filtering technique also begins with the customer, but this time evaluates the range of products that the customer tends to select and compares that with the aggregation of the product profiles of other customers that tend to purchase similar items.

	Snickers	CocoPuffs	Tomato	Avocado
Customer 1	4x/week	1x/week	⊘	⊘
Customer 2	2x/week	⊘	4x/week	2x/week
Customer 3	⊘	1x/week	3x/week	4x/week
Customer 4	2x/week	1x/week	⊘	?

Table 1: Collaborative Filtering Example

For example, let's take a look at Table 1. If we want to figure out if an avocado is valuable to Customer 4, then we can look to the product preferences of Customers 1-3 who were identified as similar users because of their shared preferences for Snickers, CocoPuffs and Tomatoes. We cannot determine an outcome from Customer 1's profile because of the lack of information regarding the Avocado. We can, however, determine that Customer 4 is likely to enjoy an avocado based upon their shared preferences for Snickers and CocoPuffs and the additional preference of Customers 2 and 3 for avocados.

There are two techniques that are commonly used to determine which products to recommend; memory-based and model based. Within model based approaches there are a number of algorithms that are employed, including: machine learning, algebraic technology, k-means clustering, Gibbs sampling, Bayesian models, probabilistic relational models, linear regression, max entropy and more. There is great discussion surrounding which method is better to use, but in the interest of brevity and clarification, we will stick to the memory-based approach. As we continue to develop this software the model is likely to change to improve accuracy and validity [11].

The memory-based algorithm determines the value (r) of a product (s) to a customer (c) based upon the aggregate (aggr) of the value of the product to all other similar customers (c'). (K) is used as a normalizing constant, ($\bar{r}$) represents the average value of all products for the specified customer, and (sim) is a function determining the similarity between two customers.

$$r_{c,s} = \underset{c' \in \hat{C}}{\mathrm{aggr}}\, r_{c',s}$$

$$r_{c,s} = \bar{r}_c + k \sum_{c' \in \hat{C}} sim(c,c') \times (r_{c',s} - \bar{r}_{c'})$$

$$sim(x,y) = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{||\vec{x}||_2 \times ||\vec{y}||_2} = \frac{\sum\limits_{s \in S_{xy}} r_{x,s} r_{y,s}}{\sqrt{\sum\limits_{s \in S_{xy}} r_{x,s}^2}\sqrt{\sum\limits_{s \in S_{xy}} r_{y,s}^2}} \quad [11]$$

Both content and collaborative filtering techniques provide astonishingly accurate predictive recommendations for customers. As it turns out, the combination of both methods into a hybrid model can produce even more accurate results [11]. For this project we will use a weighted combination model that gives certain weight to each predicted value produced by each filtering system. The details of this design can be seen in Figure 2.

Overall Nutritional Quality Index (ONQI) Algorithm
The ingenuity of this project comes from the addition of an algorithm designed by the company, NuVal:

> The Overall Nutritional Quality Index (ONQI) is an algorithm designed to generate a single, summative score for the "overall nutritional quality" of a food based on its micronutrient and macronutrient composition and several other of its nutritional properties (e.g., energy density). The ONQI is further designed to stratify foods into a rank order of relative nutritiousness both universally (i.e., across all food categories) and within specific food categories (e.g., breads, cereals, frozen desserts, etc.), while avoiding the characterization of any food as "good" or "bad" in absolute terms [12].

The ONQI produces what is called a trajectory score, which is a value that falls within the range of 0 to 100, 100 being

the healthiest option. For each item the score is determined by dividing the nutrients with favorable effects on health by the nutrients with unfavorable effects on health, weighting each nutrient by its' impact on public health. In Table 2 and Figure 1 you can see the list of nutrients under consideration and the algorithm that is used to determine the trajectory score [12].

Numerator Nutrients*	Denominator Nutrients*	Universal Adjustors*
Fiber	Saturated fat	Fat Quality
Folate	Trans fat	Protein Quality
Vitamin A	Sodium	Energy Density
Vitamin C	Sugar (total/added)	Glycemic Load
Vitamin D	Cholesterol	
Vitamin E		
Vitamin B12		
Vitamin B6		
Potassium		
Calcium		
Zinc		
Omega-3 fatty acids		
Total bioflavanoids		
Total carotenoids		
Magnesium		
Iron		

*NOTATIONS WITHHELD.

Table 2: Nutrients included in ONQI algorithm [12]

1st Universal Adjustor×2nd Universal Adjustor×Sum of Weighted log transformed TS of Numerator Nutrients

Glycemic Load×Energy Density×Sum of Weighted log transformed TS of Denominator Nutrients

Or

$(1+UA1)\times(1+UA2)\times(1+W_{P1}\times W_{S1}\times W_{R1}\times\log(1+TS_1)+ - - - +W_{P16}\times W_{S16}\times W_{R16}\times\log(1+TS_{16}))$

$GL\times ED\times(1+W_{P1}\times W_{S1}\times W_{R1}\times\log(1+TS_1)+ - - - +W_{P5}\times W_{S5}\times W_{R5}\times\log(1+TS_5))$

Figure 1: ONQI Formula [13]

This trajectory score completes our dynamic coupon system by acting as a sorter of the recommended products. Instead of receiving coupons for the items that they prefer the most, customers will receive coupons for the items that they prefer the most and that are also healthier for them. With over 30,000 food items already registered in the ONQI database, this dynamic coupon model is already capable of being implemented.
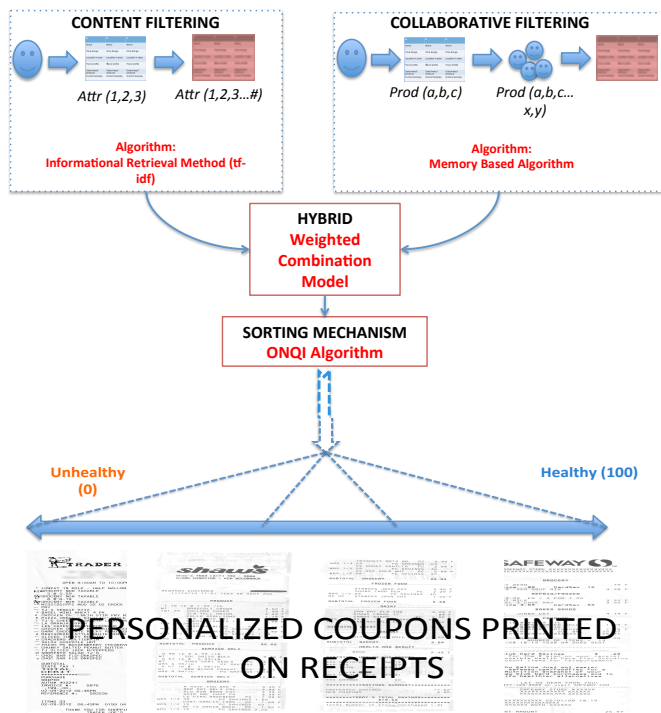
Figure 2: Smart^{Healthy} Coupon Model

Companies like Catalina Marketing and Inmar are two of many companies that are perfecting the most effective filtering methods for dynamic coupons. Smart^{Healthy} Coupons are different because it targets the populations that utilize coupons the most and incorporates meaningful corporate social responsibility by nudging users towards healthier products.

The benefit of this approach is that as a form of intervention, it does not require the individual to radically shift their behavior within a short time frame. In fact, this intervention borrows from B.J. Fogg's Tiny Habit Theory that purports that individuals can change their overall behavior if they start by making small changes. Smart^{Healthy} Coupons aims to fulfill individual food preferences with superior options.

USER SCENARIO
Smart^{Healthy} Coupons serves the customer by becoming integrated in his/her local grocery store at which he/she has a membership. When a customer checks out, they will receive various coupons printed at the end of their receipt. The inspiration of using the receipt as a vessel for the next set of coupons comes from issues in accessibility for low-

income populations. A receipt that can be printed online is a lot more difficult for someone who does not have home access to a computer or cannot afford a smart phone.

The customer will receive the coupons on the back or at the bottom of their receipt and will be able to use it within a given time period. These coupons will offer products that are slightly healthier than their current purchases. If the customer uses these coupons with high frequency, the NuVal system will sort towards healthier and healthier products, again, within the range of preferences of the individual.

The type of products that will be presented can be understood by understanding the results of each filtering mechanism with the ONQI sorter. The content filter will produce products that are similar to current purchases, but when sorted by the ONQI, will recommend items that are almost the same but healthier. For instance, if the individual purchases spaghetti, the system is likely to recommend whole wheat spaghetti due to its higher fiber content. The collaborative filter will produce products that other similar users frequently purchase, but when sorted by the ONQI, will recommend healthier items that these similar users purchase. For example, these users who share the target customer's preference for snickers, might also purchase broccoli and icecream. The ONQI sorter will give priority to the broccoli because it will yield a higher trajectory score. In light of the hybrid model, users will receive coupons for products that are determined in both ways.

DESIGN DECISIONS
The hope and expectation from this project is that it will not drastically change the current profit driven environment for the manufacturer, retailer and individual grocery store. It is true that users will be nudged towards healthier options, however, many of these manufacturers now carry healthier alternatives which can supplement any decrease of revenue that may come from individuals purchasing less unhealthy food. Additionally, this recommender system is still highly tailored to individual preferences so the hope is that users will respond to these coupons in a very similar manner. If we can demonstrate that the return on the coupons is no different from the standard dynamic coupon model, then our service provides an added bonus of high social corporate responsibility and welcomed public relations. This would allow Smart^{Healthy} Coupons become a strong competitor in the dynamic coupon market.

We will need to develop applications at the level of the manufacturer, the retailer, and the store in order to ease the transmission of data and cooperation. These applications will facilitate all economic decisions based upon customer behavior.

EVALUATION
The process of evaluation can be identified in three categories: algorithms, behavior, and health outcomes.

Algorithms

In order to evaluate the effectiveness of the algorithms involved we can calculate the mean squared error and/or the root squared error. This will determine how accurate the algorithms may be and whether or not the recommendations they produce are valid according to the model.

Behavior

The behavior of the customers is the most important measurement for conducting business with the manufacturers, retailer and stores. We will need to identify if the individual used the coupon, how quickly the coupon was used, whether the individual continued to purchase the item even without the coupon and whether or not the individual was inspired to purchase more products than usual because of the coupon. All of these factors will determine the economic benefit of imploring these coupons.

Health Outcomes

From a public health perspective, the last and most important outcome to measure is whether these Smart[Healthy] Coupons have a positive impact on health. This evaluation will be more costly and time consuming but can be achieved through the implementation of a randomized control trial. By comparing two grocery stores of the same retailer that receive similar demographics in customers, we are able to implement our coupon system in one store and not the other. We can offer incentives for customers at both stores to sign up to be apart of our study and we can track their health status over time. The hope is that we will see a decrease, or at the very least, a slower growth rate of obesity and obesity related illnesses in the store that receives our coupons.

LIMITATIONS

A key issue facing Smart[Healthy] Coupons is the emphatic opposition towards price discrimination by the public. Although this recommending customer specific coupons is not price discrimination in the classical sense, it may be possible that users take offense to not being offered the same discounts for the same products. Hopefully, companies like Catalina Marketing and Inmar have found a way to make that work. With a bit of investigation we can determine whether dynamic coupons do not bother customers or whether these other companies have found a work around.

There are other possible legal implications in the recommendation of the products. A concern for the health safety implications of the product that gets recommended to the individual such as a product that uses a sugar alternative that does not have enough peer-reviewed information to understand its effect on health. The rebuttal to this claim is that NuVal continuously updates its algorithm in light of this information. If there is not conclusive evidence regarding the primary ingredients within the product, then that product will not be selected until their committee of specialists have deemed it safe.

Another issues is customer privacy since these individuals signed up for an account with the retailer, they do not expect to have their data shared with third party entities. A solution would be for Smart[Healthy] Coupons to never take ownership of customer data. Instead, this would be a software that can be adapted and integrated into the cataloging system in place at the manufacturer, retailer and store. That is why we specified before the necessity of applications designed for the manufacturer, retailer and store.

The final issue is that of a technological nature. Many of these algorithms implored with Smart[Healthy] Coupons face obstacles by their very design. Content filtering requires assigning many of these attributes for all products by hand. Additionally, it is difficult for a content filter to distinguish between products with the exact same attributes. In addition, both content and collaborative filtering face barriers because without enough data on customers and products to begin, the algorithms cannot be effective. Fortunately, if we can partner with retailers that already keep a digital log of much of this information, we are less likely to run into these problems [11].

WHAT DID I LEARN?

The most difficult challenge I faced was identifying the correct vocabulary necessary to conduct relevant research. I am applying skills I have learned in public health to a whole new field within computer science, and understanding how those fit together was quite difficult. Meetings with professionals such as Pattie Maes and Erik Brynjolfsson were integral by inspiring a foundation of vocabulary to begin the search. From there I utilized social websites such as Quora and Reddit to aid in my search for relevant material. Once I was able to identify a few articles I used the citations as gateways to the most relevant information for my project. This process was by and large the most difficult and time-consuming portion of this project, as I had to navigate an entirely new discipline and teach the concepts to myself.

I found this process to be highly engaging and incredibly beneficial and, in fact, prefer this form of learning than most. The process of discovering the correct material felt a bit like I was a detective for knowledge and this continuously spurned my curiosity. Once I found the information I knew it was up to me to read it, digest it and apply it to the content of my project. I learned how recommender systems work and the various techniques that are used. I was able to draw upon my background in physics to understand the formulas of the algorithms. Additionally, I learned how to leverage resources online to inspire and inform my work. By analyzing NuVal's reference manuals, I was able to spot a small notation regarding the patent pending status of their product. By using the Google patent search, I was able to identify their patent and subsequently extract their entire formula for their trajectory score (ONQI algorithm). This entire process

taught me how to critically search as well as teach myself aspects of a new discipline.

FUTURE WORK

My immediate plan going forward is to create a proposal and submit it to Erik Brynjolfsson's fall class as a possible case for a consulting project. Brynjolfsson has designed a course for Sloan students to consult on various cases and if mine is selected, students will design a business proposal over the course of a semester. Concurrently, I will work with a few friends who have experience in software design to draft a working model of this system to demonstrate how it functions. Considering the progress and the feedback I receive along the way, I plan to either implement it myself or consult with Catalina Marketing or Inmar to propose their adoption of this model and subsequent partnership with NuVal.

ACKNOWLEDGMENTS

REFERENCES

1. Everson, S.A., et al., *Epidemiologic evidence for the relation between socioeconomic status and depression, obesity, and diabetes*. Journal of psychosomatic research, 2002. 53(4): p. 891-895.

2. Drewnowski, A., *The cost of US foods as related to their nutritive value*. The American journal of clinical nutrition, 2010. 92(5): p. 1181-1188.

3. Dammann, K. and C. Smith, *Food-related attitudes and behaviors at home, school, and restaurants: perspectives from racially diverse, urban, low-income 9- to 13-year-old children in Minnesota*. Journal of nutrition education and behavior, 2010. 42(6): p. 389-397.

4. Adam, T.C. and E.S. Epel, *Stress, eating and the reward system*. Physiology & Behavior, 2007. 91(4): p. 449-458.

5. Jabs, J. and C.M. Devine, *Time scarcity and food choices: an overview*. Appetite, 2006. 47(2): p. 196-204.

6. Maffeis, C. and L. Tato*, Long-term effects of childhood obesity on morbidity and mortality*. Hormone Research in Paediatrics, 2004. 55(Suppl. 1): p. 42-45.

7. Schwimmer, J.B., T.M. Burwinkle, and J.W. Varni, *Health-related quality of life of severely obese children and adolescents*. Jama, 2003. 289(14): p. 1813-1819.

8. Ward, C., *How Google's Algorithm Silences Minority Opinions*, in Highspeed Internet. 2014.

9. Dong, D. and H.M. Kaiser, *Investigating household food interpurchase behavior through market segmentation*. Agribusiness, 2010. 26(3): p. 389-404.

10. LaVeist, T.A., D.J. Gaskin, and P. Richard, *The economic burden of health inequalities in the United States*. 2009.

11. Adomavicius, G. and A. Tuzhilin, *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*. IEEE Trans. on Knowl. and Data Eng., 2005. 17(6): p. 734-749.

12. David L. Katz, M., MPH, et al., *Overall Nutritional Quality Index: Version 1 (Abbreviated Reference Manual)*, G. Hospital, Editor. 2007, Prevention Research Center: Yale University School of Medicine: Derby, CT.

13. Katz, D.L., *Method and system for scoring a diet*. 2013, Google Patents.