Layout Optimization and Promotional Strategies Design in a Retail Store based on a Market Basket Analysis

Conference Paper · July 2016

DOI: 10.18687/LACCEI2016.1.1.307

3 authors, including:

Kevin Apolinario
Escuela Superior Politécnica del Litoral (ESPOL)

Andres G. Abad
Escuela Superior Politécnica del Litoral (ESPOL)

Layout Optimization and Promotional Strategies Design in a Retail Store based on a Market Basket Analysis

Ing. Jonathan Bermudez, Ing. Kevin Apolinario, and Andres G. Abad, Ph.D.
Escuela Superior Politecnica del Litoral (ESPOL), Guayaquil, Ecuador,
{jojaberm, kevraapo, agabad}@espol.edu.ec

Abstract* – Retail transactional data is a rich source of information offering new dimensions for company's competition. The challenge is to turn data into useful insights with business value. One widely used technique for analyzing transactional data is Market Basket Analysis (MBA), which is a data mining technique that studies what customers' are buying together. Thus, MBA is a contingency study and not a correlation analysis. In this paper we present a case study for analyzing retail transactional data using MBA and use the results as a prescriptive model for sell floor optimal design and for guiding the design of related in-store promotional strategies.

***Keywords*– Market Basket Analysis, Affinity Analysis, Sell Floor Layout Optimization, In-store Promotional Strategies.**

I. Introduction

Retail transactional data is a rich source of information generated at points of sale (POS). The volume and velocity of this data is usually so high that traditional data analysis techniques fail. In fact, POS transactional data is usually referred to as a canonical instance of Big Data [1]. Additionally, analysis of transactional data rarely includes anything else aside from descriptive statistics; predictive and prescriptive analysis are rarely performed.

Turning POS records into useful business insights is an attractive endeavor for retail firms. In particular, two ways in which POS records may be used are (1) as input in the sell floor layout design, and (2) in designing in-store promotional strategies based on observed purchasing behavior.

It has been widely recognized that sell floor layout plays a crucial role in customers' experience at retail stores. For example, it was observed in [2] that "selling floor layout are extremely important because they strongly influence in-store traffic patterns, shopping atmosphere, shopping behavior, and operational efficiency." In [3] the authors relate store layout design with customers' price acceptability and their willingness to pay, while in [4] the authors establish a relationship with customers' preferences and interests. Store layout design is identified as a determining factor of store loyalty in [5].

On the other hand, firms design promotional strategies to convince their customers—and potential new customers—to spend more at their stores. Promotional strategies are particularly important for retail firms competing in competitive markets, because of their need to achieve high-volume sales at low-margin rates.

Data mining is used to turn data into useful valuable knowledge. Its application in the retail industry is quite broad, ranging from micro-segmentation of customers [6], churning prediction and customers' retaining [7], inventory management [8], price optimization [9], and customer's sentiment analysis [10].

Affinity analysis is a data mining technique in which the co-occurrence of events of interests are studied [11]. Market basket analysis (MBA) is a particular type of affinity analysis in which the events of interest are the buying of products [12]. This type of analysis has been popularized by successful e-commerce applications such as amazon's *"Customers who bought this item also bought..."* feature. In general, MBA quantifies the *complementary* and *supplementary* relationship between products.

In this paper we present a case study that uses MBA on POS records to attain two objectives: (1) to provide a prescriptive model for store layout optimization, and (2) to design promotional strategies for up-selling and cross-selling. The case study was conducted at a particular store from a supermarket retail chain in Latin America.

The two objectives can benefit from the results of an MBA. To see this, consider two complementary products, as determined by an MBA. We could use this information to define a sell floor layout to *induce* clients to increase their in-store walking distance, and, also, to direct particular promotional campaign to try to enforce that customers keep buying them together. The relationship between MBA and the two objectives is schematized in Figure 1.

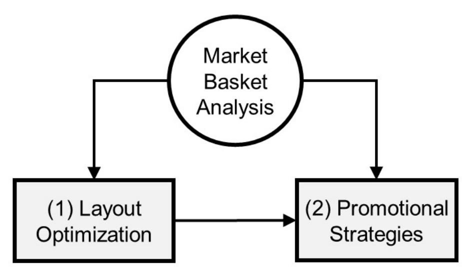

Fig. 1: Relation between an MBA and the two objectives.

Layout optimization have a direct effect on the design of promotional strategies. We consider the design of in-store promotional strategies based on in-store advertising and on a looking-and-reward model [13]. Specifically, we placed ads inside the store, next to particular products. The ads promote complementary products. Thus, this marketing strategy interacts with the layout optimization. Note that the relationship is one-way, since the campaign does not determines the optimal layout.

The rest of the paper is organized as follows. In Section II we conduct some knowledge discovery processes including basic descriptive statistics on the number of distinct product families per tickets, and we describe the MBA and present some resulting association rules. In Section III we describe the consolidation of the knowledge obtained to define an optimal sell floor layout and to design in-store promotional strategies. Section IV discusses the evaluation of implementing the proposed solutions. Section V concludes the paper providing a discussion and offering future directions to exploit.

II. Discovering Knowledge

Every successful data mining application begins with a well conducted Knowledge Discovery process. During this stage, data is analyzed—in usually creative and novel ways—resulting in the discovery of new useful insights. Here we present a basic exploratory data analysis centered mainly at describing the diversity of product families within sell tickets. We then perform an MBA on the dataset and obtain the association rules that will serve as input for attaining the case study's objectives.

A. Exploratory Data Analysis

We consider transactional data corresponding to a year of operations at one particular retail store, consisting of approximately 19 thousand tickets. The store considered has been operating steadily over the last three years. In particular, the store has not been affected by external factors during the year of study and no new competitors were introduced nearby during this time.

Based on the company's practice, products are grouped in around a hundred product families. Figure 2 shows a histogram of the number of distinct product families contained in each ticket. On average, sell tickets contain 28.37 distinct product families, while the median is 26 product families. In this work we focus on the 24 product families with a higher occurrence in sale tickets. Figure 3 shows the amount of tickets containing each of these 24 product families.

To determine which products are being bought together we next describe a co-occurrence analysis.

B. Market Basket Analysis

In this section we study what customer's are buying together. For this we conduct an MBA, which performs an analysis of the co-occurrence of the buying of products. We clarify that MBA is not a correlation analysis but a contingency study.

Market basket analysis offers a way to systematically characterize customer's behavior based on hard data, offering great potential for operations management throughout the retail store. However, besides applications related to up-selling and cross-selling, MBA has not been broadly applied.

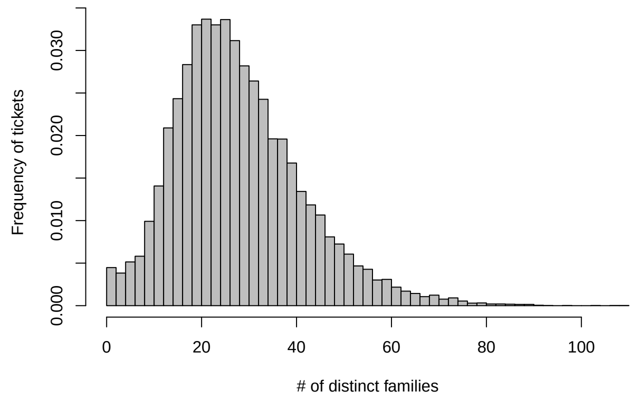

Fig. 2: Number of distinct product families contained in each ticket.

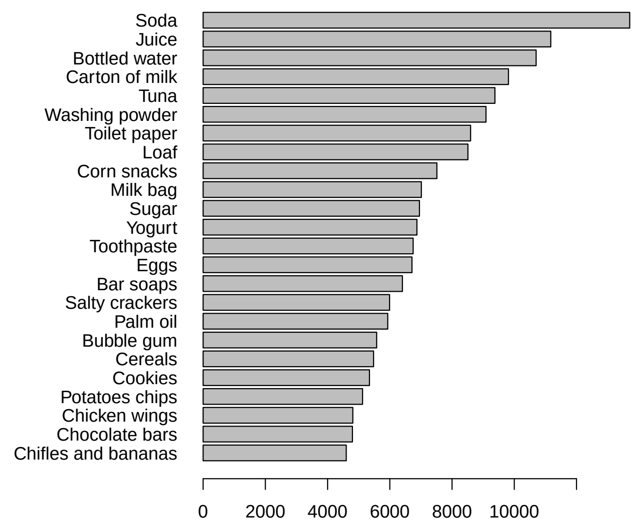

Fig. 3: Number of tickets containing at least on item from each of the 24 product families considered.

The output of an MBA is a set of association rules defined over itemsets. An itemset is defined as a set of items bought together. Association rules are, then, statements of the form

$$\mathcal{X} \Rightarrow \mathcal{Y}, \tag{1}$$

which reads: *customer who bought itemset $\mathcal{X}$ also bought itemset $\mathcal{Y}$*. Here, $\mathcal{X}$ is usually referred to as the *antecedent*, while $\mathcal{Y}$ is referred to as the *consequent*.

Even for a moderate number of items, the number of possible rules is enormous. Thus, there is interest in developing efficient algorithms for performing MBA. Most of the recent related work dealing with large instances of this problem has focused on distributed computing paradigms, such as MapReduce [14]–[16]. Most of these methods are based on the Apriori algorithm [17], a remarkable method for efficiently performing MBA. The algorithm works by applying specific selection criteria for dealing with the large number of possible rules.

The criteria used are based on the *support* of an (antecedent) itemset, and the *confidence* and *lift* of a rule. The support $S(\mathcal{X})$ of an itemset $\mathcal{X}$ is defined as

$$S(\mathcal{X}) = \mathbb{P}(\mathcal{X}). \tag{2}$$

Rules with a low support of their antecedent occur very rarely. The confidence $C(\mathcal{X} \Rightarrow \mathcal{Y})$ of a rule $\mathcal{X} \Rightarrow \mathcal{Y}$ is given by

$$C(\mathcal{X} \Rightarrow \mathcal{Y}) = \mathbb{P}(\mathcal{Y}|\mathcal{X}) = \frac{\mathbb{P}(\mathcal{X} \cap \mathcal{Y})}{\mathbb{P}(\mathcal{X})}. \tag{3}$$

Confidence may be used to determine complementary (high confidence) and supplementary (low confidence) products. Finally, the lift $L(\mathcal{X} \Rightarrow \mathcal{Y})$ of a rule $\mathcal{X} \Rightarrow \mathcal{Y}$ is given by

$$L(\mathcal{X} \Rightarrow \mathcal{Y}) = \frac{\mathbb{P}(\mathcal{X} \cap \mathcal{Y})}{\mathbb{P}(\mathcal{X})\mathbb{P}(\mathcal{Y})}, \tag{4}$$

and can be seen as a measure of the degree to which the itemsets in a rule departs from independence. In these definitions the probability concept is given a frequentist interpretation.

From the large list of possible association rules obtained from the data, we considered rules with a lift of at least 1. Overall, the MBA provides us with about 60 association rules of interest. Table I list the top ten associations rules obtained. Additionally, some trivial rules and some rules from which it is difficulty to obtain profit are also shown in the table. As an example of a rule with a high confidence and belonging to the latter group consider toilet paper and tuna: it can be difficult to promote them together because these products are not naturally related.

TABLE I: Obtained association rules

Top rules

Antecedent	Consequent	Support	Confidence	Lift
Potatoes Chips	Soda	0.205	0.705	1.086
Loaf	Juice	0.203	0.728	1.228
Eggs	Cartoon Milk	0.254	0.728	1.259
Loaf	Tuna	0.250	0.600	1.180
Corn snacks	Soda	0.273	0.709	1.092
Loaf	Cartoon Milk	0.250	0.560	1.350
Soda	Juice	0.250	0.490	1.170
Sugar	Juice	0.250	0.560	1.140
Bath Soap	Toilet Paper	0.250	0.570	1.240
Juice	Bottled Water	0.250	0.530	1.160

Trivial rules

Antecedent	Consequent	Support	Confidence	Lift
Cereal	Cartoon Milk	0.212	0.708	1.225
Toilet Paper	Detergent	0.250	0.610	1.240

Difficulty-to-exploit rules

Antecedent	Consequent	Support	Confidence	Lift
Juice	Airtime	0.212	0.708	1.225
Toilet Paper	Tuna	0.250	0.800	1.170

III. Consolidating Discovered Knowledge

After a knowledge discovery process, every effective data mining application must consolidate the discovered knowledge in practical, feasible, and beneficial ways. Here, we use the obtained association rules to define an optimal sell floor layout and to design in-store promotional strategies. The layout obtained is focused on locating complementary products far away, while locating supplementary products nearby. This is based on the assumptions that increasing the in-store travel time increases unplanned purchases [22], [23]. To support customer's decision to take longer trips across the sell floor in search of particular products we propose to use in-store ads offering discounts for purchasing complementary products.

A. Sell Floor Layout Optimization

Traditionally, sell floor layout is determined based on the store manager's own expertise. Products were distributed across the sell floor mainly based on their functional similarities. While this criterion may be effective for reducing search time and, perhaps, customer's cognitive burden; it does not take advantage of factual customer's purchase behavior obtained from historical data.

Note that the scope of our method does not include product assortment or shelf space allocation. For a data mining treatment of these problems see, for example, [18].

A full store layout optimization problem is highly challenging due to the exponential growth of the size of the solution space. In fact, the problem is $\mathcal{NP}$ hard and, thus, its solution is impractical for larger than moderate size problems. For this reason, approximate solutions are of interest. Simple heuristics have been used to solve the problem [19], [20]. A refined heuristic approach based on a network flow model was proposed in [21].

Here we propose to determine a sell floor layout configuration based on the association rules obtained in Section II-B. These rules were obtained on the basis of the 24 product families listed in Section II-A. Figure 4 shows the original layout; lines shown correspond to product families with large confidence. Note the concentration of highly related product families at the back of the store.

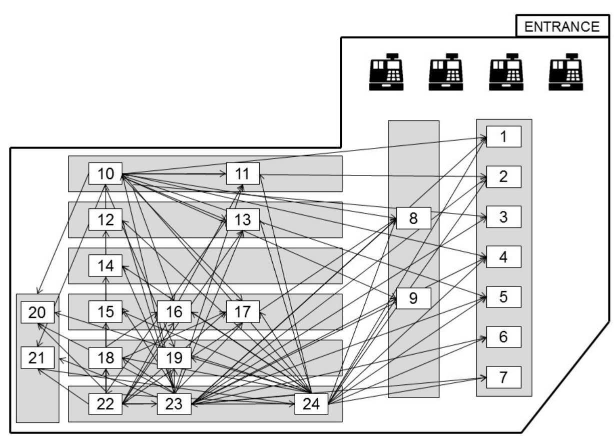

Fig. 4: Original sell floor layout.

In recent works, unplanned spending at retail stores has been related to within-trips and store travel distances [22], [23]. In this spirit, we propose to define the sell floor layout as the solution to an optimization problem aimed at maximizing the total travel distance.

Let d_{ij} denote the distance between pair of products i and j. Each pair of products is weighted by their individual prices

p_i and p_j, and by the confidence c_{ij} obtained from equation (3) as

$$c_{ij} \triangleq C(i \Rightarrow j). \tag{5}$$

The variables are

$$x_{il} = \begin{cases} 1 & \text{if product } i \text{ is located at} \\ & \text{position } l, \\ 0 & \text{else.} \end{cases}$$

The problem may be then formulated as an assignment problem with a quadratic objective of the form

$$\sum_{ijlk} x_{il} x_{jk} d_{lk} c_{ij} p_i p_j. \tag{6}$$

To cast the problem as a linear, albeit sparse, formulation we introduce variables $y_{ijlk} \equiv x_{il} \cdot x_{jk}$, defined as

$$y_{ijlk} = \begin{cases} 1 & \text{if products } i, j \text{ are located at} \\ & \text{positions } l, k \text{, respectively,} \\ 0 & \text{else.} \end{cases}$$

Using these parameters and variables, we propose the following IP problem

$$\max \quad \sum_{ijlk} y_{ijlk} d_{lk} c_{ij} p_i p_j \tag{7}$$

$$s.t. \quad 2y_{ijlk} \leq x_{ik} + x_{jl}, \qquad \forall i, j, l, k \tag{8}$$

$$\sum_i x_{ik} = 1, \qquad \forall k \tag{9}$$

$$\sum_k x_{ik} = 1, \qquad \forall i \tag{10}$$

$$y_{ijlk} \in \{0, 1\}, x_{ik} \in \{0, 1\}. \tag{11}$$

The problem was solved using IBM ILOG CPLEX Optimizer version 12.6.3.

The solution to optimization problem (7-11) is shown in Figure 5, and achieves a total distance of 1651.94 m. The original total distance was 1388.05 m, so the total distance increment is 263.89 m. As expected, products with larger confidence are located further away from each other. A similar situation is observed with respect to price: the most expensive products get further apart, while products with lower price get closer together.

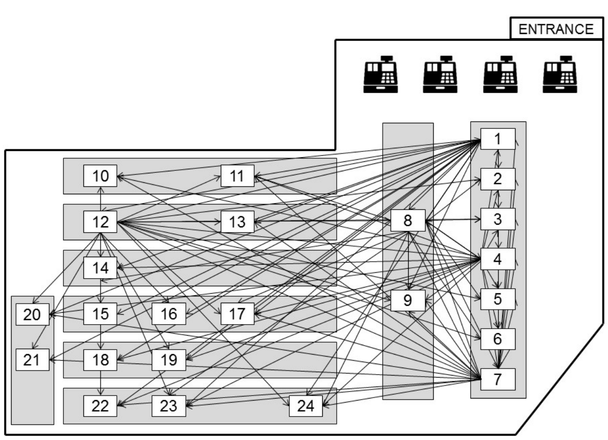

Fig. 5: Optimized sell floor layout.

B. In-store Promotional Strategies

We propose to define in-store promotional strategies to support the increment of customer's walking distance induced by the optimal sell floor layout. In particular, for complementary product families, we place an ad next to antecedents product families promoting their consequent.

For example, consider an association rule by which customers who buy cereal will probably buy milk as well. The optimal layout distribution would have probably located cereal and milk far apart from each other. Then, customers need a sense of reward to support the decision of walking from cereal to milk. In-store ads provide such an incentive by announcing associated discounts—for this particular example there will be a discount offer for 6 bottles of milk, located next to cereals. Promotions could be used similarly for up-selling campaigns.

IV. Expected Revenue Increment

The most direct approach to evaluate the effect of the optimal layout distribution and the application of the promotional strategies is to implement it in a real store. However, this incurs in high financial, time, and other resource costs. Furthermore, retail firm's policies most surely discourage arbitrary store experimentation.

To evaluate the proposed solutions, we propose the following arguments. We start by assuming that the increase of exposure of products due to longer in-store trips increases unplanned purchases. The annual value per linear meter is measured as the ratio between annual revenues and the linear length of the store. The store's annual value per linear meter is obtained as \$1,275.09. The increase of the total distance between products (263.89 m) can be then considered to have a direct proportional effect of \$336,483.50 for the store.

V. Conclusions and Future Directions

In this paper we presented an application of data mining tools to boost sells. We first performed an MBA over transactional data corresponding to the course of a year of normal operations at one store. Over 60 associations rules described the behavior of the customers. These rules were used to obtain a sell floor optimal layout. The optimized floor layout effect was supported by the design of promotional strategies gear toward justifying the increase of in-store customers' walking distance.

The association rules and the optimal layout obtained was done considering only 24 product families. Considering individual products can certainly improve the results. Approximate heuristic approaches, distributed computing implementations for performing the MBA, and large-scale optimization methods for determining the layout problem should be explored. Furthermore, the analysis could consider loyalty or reward cards which add a new dimension to the analysis, allowing for promotional strategies to be designed individually for each customer.

Acknowledgment

The authors would like to thank the Board of Directors of Tiendas Industriales Asociadas Sociedad Anónima (TIA S.A.), a leading supermarket retail chain in Ecuador, for authorizing their company to provide the authors with historical sales data for hundreds of their products.

REFERENCES

[1] V. Mayer-Schnberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, reprint edition ed. Boston: Eamon Dolan/Mariner Books, Mar. 2014.

[2] D. M. Lewison, *Retailing*, 6th ed. Upper Saddle River, N.J: Prentice Hall College Div, Dec. 1996.

[3] D. Grewal and J. Baker, "Do retail store environmental factors affect consumers' price acceptability? An empirical examination," *International Journal of Research in Marketing*, vol. 11, no. 2, pp. 107–115, 1994.

[4] I. Simonson, "The effect of product assortment on buyer preferences," *Journal of Retailing*, vol. 75, no. 3, pp. 347–370, 1999.

[5] B. Merrilees and D. Miller, "Superstore interactivity: a new self-service paradigm of retail service?" *International Journal of Retail & Distribution Management*, vol. 29, no. 8, pp. 379–389, 2001.

[6] M. Wedel and W. A. Kamakura, *Market segmentation: Conceptual and methodological foundations*. Springer Science & Business Media, 2012, vol. 8.

[7] J. Ranjan and V. Bhatnagar, "A review of data mining tools in customer relationship management," *Journal of Knowledge Management Practice*, vol. 9, no. 1, 2008.

[8] P. K. Bala, "Data mining for retail inventory management," in *Advances in Electrical Engineering and Computational Science*. Springer, 2009, pp. 587–598.

[9] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers, "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, Tech. Rep., May 2011.

[10] K. Floyd, R. Freling, S. Alhoqail, H. Y. Cho, and T. Freling, "How online product reviews affect retail sales: A meta-analysis," *Journal of Retailing*, vol. 90, no. 2, pp. 217–232, 2014.

[11] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, 2nd ed. New York, NY: Springer, Apr. 2011.

[12] M. Dhanabhakyam and M. Punithavalli, "A survey on data mining algorithm for market basket analysis," *Global Journal of Computer Science and Technology*, vol. 11, no. 11, 2011.

[13] G. S. Linoff and M. J. A. Berry, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 3rd ed. Indianapolis, IN: Wiley, Apr. 2011.

[14] J. Woo and Y. Xu, "Market basket analysis algorithm with Map/Reduce of cloud computing," in *The 2011 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2011), Las Vegas*, 2011.

[15] J. Woo, "Apriori-Map/Reduce Algorithm," in *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA)*. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012, p. 1.

[16] N. Li, L. Zeng, Q. He, and Z. Shi, "Parallel implementation of apriori algorithm based on MapReduce," in *Software Engineering, Artificial Intelligence, Networking and Parallel & Distributed Computing (SNPD), 2012 13th ACIS International Conference on*. IEEE, 2012, pp. 236–241.

[17] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A. I. Verkamo, and others, "Fast Discovery of Association Rules." *Advances in knowledge discovery and data mining*, vol. 12, no. 1, pp. 307–328, 1996.

[18] M.-C. Chen and C.-P. Lin, "A data mining approach to product assortment and shelf space allocation," *Expert Systems with Applications*, vol. 32, no. 4, pp. 976–986, 2007.

[19] F. S. Zufryden, "A dynamic programming approach for product selection and supermarket shelf-space allocation," *Journal of the operational research society*, pp. 413–422, 1986.

[20] M.-H. Yang, "An efficient algorithm to allocate shelf space," *European journal of operational research*, vol. 131, no. 1, pp. 107–118, 2001.

[21] A. Lim, B. Rodrigues, and X. Zhang, "Metaheuristics with local search techniques for retail shelf-space optimization," *Management Science*, vol. 50, no. 1, pp. 117–131, 2004.

[22] S. K. Hui, J. J. Inman, Y. Huang, and J. Suher, "The effect of in-store travel distance on unplanned spending: Applications to mobile promotion strategies," *Journal of Marketing*, vol. 77, no. 2, pp. 1–16, 2013.

[23] T. J. Gilbride, J. J. Inman, and K. M. Stilley, "The role of within-trip dynamics in unplanned versus planned purchase behavior," *Journal of Marketing*, vol. 79, no. 3, pp. 57–73, 2015.

[24] B. R. Berman and J. R. Evans, *Retail Management: A Strategic Approach*, 12th ed. Boston: Prentice Hall, Jul. 2012.