Value of Data over time in Albert Heijn’s Loyalty Program
Sander van den Bosch
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
s.f.vandenbosch@student.utwente.nl

1. INTRODUCTION

The trade-off between the usage of data on the one hand, and privacy on the other, is a hot topic in the database research field. Currently, a lot of investigation is undertaken in respect to the relation between data storage and the related decreasing privacy. The Electronic Health Record (EPD) [25], for example, recently intensified the discussion in the Netherlands. The EPD provides a complete summary of the health and medical history of a patient by gathering data from various sources and making this information accessible through an intranet to anyone who has the necessary electronic credentials to view the information. To do so, the system stores a lot of information which could violate the privacy of the data donors (the patients) when this information is accessed by the wrong people [28, 5, 6, 17].

In order to take away any misunderstandings between data users and data donors, the trade-off between the usefulness of data and privacy must become measurable. An important element in this trade-off is the value of data for the storage company (the usefulness), which we will call the service provider’s interest. On the other hand is the privacy, which can be indicated as 1 divided by the risk of misused information, as stated by Van Heerde [32]. The privacy is called the users’s interest, because it is risky for the user to have data stored at the service provider, as this data might be disclosed. If the usefulness of data can be rated (in terms of monetary worth), we can combine this with the privacy formula (indicated as 1 divided by the risk [32]) and compute the common interest for both user and service provider in each different case. By maximizing this common interest, the privacy of the data donor won’t be invaded or violated too much, but the data is still useful to the service provider, which is the ideal case in practice. Van Heerde does tell us how to measure privacy, but does not tell us how to value data. Van Heerde assumes the usefulness of data decreases over time, similar to a non-negative monotonic descending function [32], in his calculation he used a cosine. The valuing of data differs over different cases, and as a step towards a more general theory, we choose a case to examine this assumption: Albert Heijn’s Loyalty Card.

1.1 Albert Heijn’s Loyalty Card

Since 1970 companies gather information about their customers for various reasons, but eventually to make more profit. Initially, companies tried to measure and manage their customers’ satisfaction by taking satisfaction questionnaires. In the 80’s companies also started to measure customers’ rate of defection in order to prevent their customers from defecting. This migration managing is so important because large amounts of value are at stake [19, 8]. Due to past decade’s innovation, supermarkets – armed with loads of client data – are now able to determine the needs of their customers.

Currently, many supermarkets make an effort to meet their customers’ changing needs, due to moving or having a child. Also other changing needs like a new financial situation play a role [8]. To do so, supermarkets usually buy a large amount of data at research agencies. These research agencies take questionnaires at supermarket customers, asking them varying questions about their shopping experience or lifestyle. Supermarkets then adapt their discounts or change their arrangement of the shop to obey the recommendation of the researches [8]. Another way to identify the needs of customers is to gather this data yourself. This is exactly what Ahold does for their subsidiary grocery chain: Albert Heijn.

In 1997 Albert Heijn (AH) introduced the “Bonuskaart”, a loyalty card required to obtain discounts on weekly deals [1]. To obtain this loyalty card a customer has to fill in a form, consisting of several questions about their name, age, address, but also about the family composition,
and – previously – about their profession and income. Due to the "Bonusskaart", AH is able to create a link between customers and purchases. By analyzing these links in a marketing information system, AH is able to offer some special services to their customers like a personalized offer, an automatically filled shopping list, and more as we will explain later.

1.2 Risk

The analyzed data could be useful for both company and customer, respectively the data user and the data donor. On the other hand is the risk of the data being abused, or in some or another way disclosed. Due to several privacy failures from both government and participating companies[33] [7], the data storage industry is under discussion. A good example is the recent introduction of the OV-Chipkaart, a payment-system for public transportation. When the Radboud University of Nijmegen revealed several security issues [12], Trans Link Systems (manufacturer of the OV-Chipkaart) told the press the risks were not as high as the hackers stated [30]. Due to the reaction of Trans Link Systems, the travellers started to worry about their privacy.

Although AH’s loyalty card has not yet been under a large public debate, this is only a matter of time and therefore a clear trade-off between company’s and user’s interest is essential. For that reason an increased understanding of data and its usage is required.

1.3 Problem Statement

A method should be found to rate the usefulness of data, in order to underpin the trade-off between company’s and user’s interest. Van Heerde assumes that the usefulness, in terms of monetary worth, of data for the service provider decreases over time, similar to a non-negative monotonic descending function [32]. This research examines this assumption for the specific case of AH’s loyalty program. Eventually we will approve or disapprove this assumption.

1.4 Research questions

The goal of this research is to develop a method to rate the monetary worth of data over time, specifically for AH’s loyalty program.

Therefore we have to investigate several subquestions:

In general:

- What methods are currently used to rate the usability of data? (Will be addressed in section 4)
- What methods do other research fields provide in valuing goods? (Will be addressed in section 4.1)

Specifically for AH’s loyalty program:

- For what applications does AH store data? (Will be addressed in section 3)
- Can we determine a resulting effect in spendings due to the AH Loyalty Card? (Will be addressed in section 4.2)
- Can we illustrate how the value of the data varies over time? (Will be addressed in section 4.3)

The answers on the general subquestions will be derived from a literature study. The subquestions specific for AH’s loyalty program will be answered by a combination of an explanatory research and literature study.

2. RELATED WORK

Some research has been undertaken to determine the effects of loyalty programs. Yuping Liu and Rong Xang, for example, examined the impact of competing loyalty programs on market share and saturation [23]. Consumer factors, like usage.goodwill level, perceived effort advantage, shopping orientation and price sensitivity play a big role in the success of the loyalty program. Kim, Shi and Srinivasan and Koppalle and Neslin examined the effect of different rewards to the loyalty program. Giving a reward in cash is much more inefficient for a company than rewarding a customer with a product of the firm, because the satisfaction level of the customer is equal, but the costs for the company are much lower. Since Albert Heijn rewards their loyal customers with discounts on their weekly deals, this forms an efficient manner [13, 18, 19, 20, 24, 36]. Kivetz and others researched the great impact of participation requirements, the management, and also the rewards on the success of loyalty program [14, 15, 16, 9, 20, 26, 34].

All mentioned researches are focused on short-term impact. Yuping also examined the effects of loyalty programs on the long-term. This research distinguishes between heavy buyers and light buyers. On the long-term, apparently, heavy buyers don’t change their spending, in contrast to the light buyers who do increase their expenses [22].

Gallagher researched the value of a management information system as perceived by its users. For this study a medium-size firm, employing 1,800 people, was selected. The firm used an Expense and Budget (EAB) System, essentially a cost accounting system. The managers of this firm were asked to value the EAB in terms of monetary value. At the time of study, 103 managers were receiving EAB reports, and they were asked to estimate the annual dollar value in response to the hypothetical question: “Assume that your company plans to eliminate all data processing and to obtain this report from another firm on an annual subscription basis. What is the maximum amount you would recommend paying for this report for your use?”. This question evaluates the usefulness of the data maintained in the EAB-system. It turned out that the answers varies by the degree of detail and were between $176,000 and $404,000. [11]

3. APPLICATIONS OF ALBERT HEIJN’S LOYALTY PROGRAM DATA

In order to determine the value of the data AH gathers, we need to determine its purposes. We distinguish between three application fields: Applications for AH, applications for the customer, and third parties applications. To rate the usefulness of the data, all these three fields have to be considered, but only those applications which would not be possible without the loyalty card, in other words: the link between purchase and customer.

For Albert Heijn:

Initially, AH saves money by not buying the data from a research agency, but by collecting data theirselves. This is an obvious but important saving. Furthermore, AH knows – by analyzing their data – the residence of their customers. Knowing the residence enables AH to extract some extra information from the customer data:

- Where is an extra supermarket required or desirable?
- Where is it interesting, from a financial point of
Furthermore, the changes in buying behavior could be determined. The changing needs of a customer may include changes in the family, like having a child or a changing income. As soon as AH determined these changes, AH could address them by offering the customer, for example, some personal deals.

For the customer (and eventually for Albert Heijn, by providing this service):

AH could offer the customer some personal deals, based on:

- Customer interest: what is bought by the customer on a regular base?
- What other items are bought by other customers, who bought the same products as the customer in question does?
- What kind of deals does the customer in question like? So we could offer him or her more of these deals?

And, as mentioned above: the customer' changing needs will be determined, so customers will get personalized deals which (hopefully) attract their attention.

For third parties:
AH sells their collected data to Agis, a Dutch health insurance company. Agis uses this data to provide a health check to their customers, and to provide a discount to customers who do not buy unhealthy products at Albert Heijn. In this case, the profit is generated by selling the data to a third party. This application is used in one method in economics for determining the value, to which we will come later.

4. METHODS FOR DETERMINING VALUE OF DATA

In order to determine the value of data in AH’s loyalty program, we should look beyond the boundaries of computer science. Economic theories offer us various economic valuing methods for determining the value of goods, including the course of that value over time.

The gathered data is as valuable as it is used by AH’s applications. AH could store loads of data, but when it isn’t used it’s worthless. We determine the value of the gathered data by analyzing its effect on the customers, resulting in an additional profit. The theory of marginal value supports this assumption [3].

4.1 Comparing methods for determining value of data

A well-known technique for economic valuing is Stated Preference [2]. This method provides us a valid and reliable way to measure the added value of AH’s loyalty card: the additional profit. Beside Stated Preference, there are several other methods for valuing in the economic theories, but they are all used for valuing the environment or problems regarding the environment. Therefore, they are not suitable in this research.

Another method to determine the value is determining how other companies value the data. Since this is only a shift of – and not a solution to – the problem, this method won’t help. Furthermore the Stated Preference offers us a pretty good view on the additional profit gained by the loyalty card. Hence, we will use the Stated Preference method to determine the value of AH’s loyalty card.

4.2 At a specific point of time

In this section we propose a method for determining the value of the data used in the AH loyalty program, by calculating the difference in average transaction value between customers with the AH loyalty card and customers without. Included in this calculation is the resulting effect (in terms of additional profit) of the applications AH practices.

We admit that the value of the applications that AH does not practice is not included in this result. However, this proposed method gives a good indication of the value of the system. Eventually, the precise value of the data is not required to indicate the course of the value of data over time.

Briefly, the approach of Stated Preference is as follows. We take two samples: one group of customers with the AH loyalty card and a group of customers without. Both groups have to contain five hundred participants at least, in order to decrease the variance and increase the statistical confidence. Furthermore, the composition of the two groups, regarding socio-cultural characteristics, must be equal. These requirements must be fulfilled to eliminate differences in life style as much as possible. Additionally, we have to define some variables to prepare our hypothesis:

- $x = \text{average transaction value of a customer with AH loyalty card}$
- $y = \text{average transaction value of a customer without AH loyalty card}$

We define our null and alternative hypothesis as follows:

$H_0 : x = y$

$H_1 : x > y$

Because we perform two independent aselect samples, we use the Student’s t-distribution to calculate whether the means are significantly different. This calculation will eventually confirm or disprove our hypothesis. X and Y are two independent aselect samples, and both sample sizes are equal (we stated $n = m = five$ hundred); it can also be assumed that the two distributions have the same variance.

The t statistic to test whether the means are different can be calculated as follows [31, 10]:

$$T = \frac{\bar{X} - \bar{Y}}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

where

- $n = \text{number of participants in group 1, as follows five hundred}$
- $m = \text{number of participants in group 2, as follows five hundred}$

According to the Student’s t-distribution, $S$ is the standard deviation as [31, 10]:

$$S^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n + m - 2}$$

For significance testing, the total degrees of freedom in this calculation is $n+m-2$. $S_X^2$ and $S_Y^2$ represent sample variances. The sample variance of a finite sample population
is calculated as follows [35]:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

We use a upper-tail probability of 5%, and as mentioned above the total degrees of freedom is approaching infinity. Accordingly, as soon as the calculated $T > 1.64$, $H_0$ will be rejected and $H_1$ will be approved, because $T > 1.64$ means customers with a AH loyalty card spend significantly more at their purchases.[27]

A second calculation is required to determine how much additional profit is gained.

We do this by drawing the profit function from Albert Heijn. Of course this is an assumption, because AH does not release this information. Before drawing the profit function, we first define some variables:

- $p =$ additional profit gained by the loyalty card
- $a =$ additional sales volume gained by a customer with a loyalty card
- $q =$ number of customers with a loyalty card
- $c =$ additional costs for customers with a loyalty card
- $s =$ savings costs customer surveys, as indicated at the applications for AH
- $e =$ a vector with all negative externalities (transaction spillover), added to the costs of the loyalty card

$e$ represents for example the people who avoid the AH due to the loyalty card. Also the costs of the system maintenance is included in $e$

The additional profit gained by the loyalty card can be calculated by the additional sales volume per customer (a) times the number of customers with a loyalty card (q). We have to subtract the additional costs for the loyalty card for AH per customer, times the number of customers ($c*q$). Furthermore, AH saves on buying loads of client data at research agencies, because they gather this data themselves, we have to add this savings to the profit function ($s$). Finally we have to subtract the vector with all negative externalities, like people who avoid AH because they think their privacy cannot be guaranteed ($e$).

The additional profit formula turns out as follows:

$$p = (a * q) - (c * q) + s - e$$

With this formula we can calculate the additional profit gained by AH due to the loyalty card. The additional profit is, as stated above, a good indicator of the value of the data, because it takes the result of the data into account. The outcome of this formula will be the input for the course of the value.

4.3 Over time

As soon as we have determined the value of some data, we then like to know what the course of this value over time is. Economic theories offer us a good method to calculate the Present Value. Present value is the value on a given date of a future payment, taking the time value of money and other factors such as investment risk into account.

As input for the Present Value formula we will take the additional profit generated by the AH loyalty card in year $t$. The course of the value over time can be calculated using the following Present Value calculation [4, 21, 29]:

$$PV = \sum_{t=0}^{N} \frac{PY_t}{(1 + i)^t}$$

Where

- $t =$ time in years
- $i =$ real interest, at companies we always take 10%, thus 0.1 in this formula
- $PY_t =$ Additional profit in year $t$

Example: We want to calculate the total value of data in the system in 2010, 2011 and 2012. We assume the additional profit gained in these years are equal, let us say: 100. By not taking into account additional profits of future years, the monetary value of the data item in 2010 follows the course shown in figure 1.

![Figure 1. Value of a data item over time.](image1)

Every future year we gain an additional profit of 100. Accordingly, the following turns out:

$$100 + 100/1.1^1 + 100/1.1^2 = 273.54$$

As we extend this calculation to the year 2054, the course of the total value of data in the loyalty system is shown in Figure 2.

![Figure 2. Total value of data items in system.](image2)

5. DISCUSSION

In the research above, we used the method of stated preference. This method, unfortunately, does not take all applications into account, because it just considers the result (in terms of additional profit) of the currently used applications. If the data would be further analyzed, new conclusions might be drawn which might result in a higher transaction value to customers with an AH loyalty card. Therefore the value of the data could be increased by further analyzing the data, and this value has not yet been taken into account.

Furthermore we have to point out an important condition of the two research groups that must be fulfilled. The preference for Albert Heijn also has to be equal, because this isn’t captured in equal lifestyles. It might be the case that customers who own a loyalty card, prefer AH over other
grocery stores, and therefore spend more at AH than customers without a loyalty card. This fact has to be taken into account in the formula. A simple but time-consuming method is to take two blank untouched research groups. Then one of them is equipped with a loyalty card. After, let us say, ten years we measure the average difference in transaction value between the two research groups. Although we cannot eliminate the effect of preference over AH, the initial preference for AH was equal.

The outcomes of this research would also be applicable in related fields. Other companies or institutions using a loyalty program could measure their success factor by using the above profit function, and also calculate the monetary value of the gathered data over time. The calculations could also be used for data not involved in a loyalty program, although a new profit function should be realized.

6. CONCLUSIONS

As an answer to the research questions stated in the introduction, the following can be concluded. As the graph presented in Figure 1 shows, the data evaluation over time does match the course of a non-negative monotonic descending function. Therefore we can underpin, based on the research above and taken the discussion into account, the assumption of Van Heerde [32]. In this case, the input for the formula was a real interest of 10%. Accordingly, the total value in the system does also match with van Heerde’s assumption, as it is an increasing function.

7. REFERENCES


