2010

THE VALUE OF ONLINE PRODUCT BUZZ IN SALES FORECASTING

Peter v.d. Reijden
Erasmus University, peter@linelink.nl

Otto Koppius
Erasmus University, okoppius@rsm.nl

Follow this and additional works at: http://aisel.aisnet.org/icis2010_submissions

Recommended Citation
http://aisel.aisnet.org/icis2010_submissions/171
THE VALUE OF ONLINE PRODUCT BUZZ IN SALES FORECASTING

Completed Research Paper

Peter v.d. Reijden
Dept. of Decision and Information Sciences
Rotterdam School of Management
Erasmus University
PO Box 1738
3000 DR Rotterdam
The Netherlands
peter@linelink.nl

Otto Koppius
Dept. of Decision and Information Sciences
Rotterdam School of Management
Erasmus University
PO Box 1738
3000 DR Rotterdam
The Netherlands
okoppius@rsm.nl

Abstract

Online product buzz refers to an online expression of interest in a product, such as online product reviews, blog posts and search trends. We study how well online buzz predicts actual sales across different phases in the product lifecycle. Using data from smartphone sales of a Dutch online retailer, we demonstrate a 28% overall increase in forecasting accuracy when measures of online product buzz variables are take into account. The value of online product buzz shows especially in early sales forecasting, an area in which traditional forecasting models have substantial difficulties. In early sales, local search trends, subscriptions for stock notifications and pageviews were important predictors and forecasts were on average improved by 44%. For mature sales, accuracy improved by 10%, with the most important predictors being on- and offsite reviews and, again, pageviews. These results also suggest different drivers of sales across phases in the product lifecycle.

Keywords: Predictive modeling, User-generated content, Web mining
Introduction

Although sales forecasting has been at the center of attention for both academics and practitioners for decades, forecasting early sales remains a largely unsolved challenge. This is striking when you take into account that in most industries – and consumer electronics in particular – the speed at which new products are introduced have increased dramatically (Chandrasekaran & Tellis, 2007), emphasizing the relative importance of early sales. In the extreme example of the movie industry, no less than 25% of total box office revenues are generated in the first weekend (Simonoff & Sparrow, 2000).

The difficulty in predicting early sales is inherent to the incredible amount of contingencies that determine a product's success – such as competition, perceived quality of the product, state of the economy, marketing efforts etc. – each of which are hard to express in numbers. It is found that the three most prominent approaches to sales forecasting have trouble coping with these contingencies, mainly due to a lack of data with predictive power. As such, the quality of a forecasting model is largely determined by the data it is based upon. Whereas historical sales data has proven to be a very reliable predictor for mature and aggregate sales patterns, no such data source has been identified for early sales forecasting.

An attempt to overcome this problem was made by Moe & Fader (2002), who found preorders to be valuable in early sales prediction. However, as this data is in many cases absent, alternative data sources are still searched for. In recent years, the hidden potential in web data is increasingly recognized. Analyzing a visitor’s trail through a website (clickstream analysis) gave way to an incredible amount of research into individual customer choice models (i.e. Bucklin & Sismeiro, 2003) and a visitor’s ‘conversion behavior’ (i.e. Moe & Fader, 2004). However, a new form of web data is emerging. The ‘socialization’ of the Web is making more and more social interactions transparent. Whereas customers initially used the Internet for information gathering only, gradually it is being used to vent opinions and converse publicly as well. Although these conversations have been taking place for ages, never before was it so easy to monitor.

This study gathers all this newly available web data under the concept online product buzz, defined as “any online expression of interest in a product”. When it concerns the upcoming launch of products, potential customers might use search engines for product information or post a message on their blog to express their curiosity. This online product buzz is being captured by a multitude of adept data aggregators that watch these conversations and monitor people’s behavior, opinions and interests. More importantly, the organizations in possession of this data are opening up their databases for everyone to use.

Online product buzz is interesting, as it provides early insight into the interest for a product. Potential customers expose themselves early by searching for a product or talking about it. The benefit of this is that whereas traditional web data (such as clickstreams) becomes available only when a potential customer is nearing the end of his buying cycle (i.e. he has landed on your product page and is checking your prices), public web data is able to capture a potential customer earlier (i.e. he mentions his excitement about an upcoming product launch on his blog). In addition to identifying the number of potential buyers, this data helps in monitoring the general word-of-mouth marketing that is partaken in (Duan, Gu, & Whinston, 2008) and the general “buzz” that surrounds a product launch.

Although often nascent, there is an increasing body of evidence that this online product buzz is indeed useful in spotting trends and improving (sales) forecasts. Urban & Hauser (2004) coined the phrase “Listening in” to describe how new combinations of customers needs could be distilled from monitoring online interactions between customers and Web-based virtual advisers. “Listening in” subtly refers to the unobtrusiveness of such monitoring: the customer will not be bothered.

The value of online buzz in forecasting has often been studied in the movie industry, as movie launches are often surrounded by a lot of buzz and early sales are extremely important. Dellarocas & Zhang (2007) found that the number of reviews posted in the first week proved highly predictive of a movie's total revenues. Mishne & Glance (2006) found how the sentiment of pre-launch reviews correlated with a movie's success and Sadikov et al. (2009) found that blog references usually preceded movie sales by a week, emphasizing the potential of buzz monitoring in early sales forecasting. But the use of online product buzz in sales forecasting is not limited to the movie industry. Gruhl et al. (2005) studied how "online chatter" could be used to predict a book's sales ranking on Amazon. Interestingly, they concluded that this was mainly helpful in predicting peaks. More recent research by Google employees into the applicability of search trend data proves promising in detecting influenza epidemics (Ginsberg,
Mohebbi, Patel, Brammer, Smolinski, & Brilliant, 2008) and improving aggregate sales forecasts in the travel, retail, automotive and home sales (Choi & Varian, 2009) as well.

Since most of past research in this area relates to cultural products such as books and films, it is an not yet known to what extent these results generalize to more regular products such as for instance consumer electronics. Furthermore, it is well-known that the reasons for buying differ across phases of the product life cycle, but what this means for the role of online buzz across the product life cycle is an open question. In short therefore, this study tests the premise that online product buzz has predictive value in sales forecasting and how this value varies across phases in the product life cycle. By doing so, this study aims to enhance the understanding of what types of online data sources are available, how valuable they are, how they should be analyzed and how one can put this into practice.

For a fine-grained analysis of the value of online product buzz in sales forecasting, its predictive power is assessed using the predictive analytics technique of random forests, at different levels of aggregation (product-specific sales versus aggregate sales for the entire product group) and in different stages of the product life cycle (introductory, embryonic, early and mature sales). In each situation the models are developed with and without online product buzz variables and their predictive accuracy is subsequently compared in a holdout sample.

Clear evidence was found that sales forecasting accuracy can be increased if measures of online product buzz variables are taken into account. On average, the results suggested a 28% improvement in predictive accuracy. As expected, the biggest improvements were shown in early sales forecasting situations. Also, results show that different types of buzz predict sales in different phases of the product life cycle.

**Literature Review**

**Sales forecasting**

The potential benefits of accurate sales forecasts are vast. When inventory is perfectly tailored to actual demand, inventory costs go down, stock outs are prevented, the risk of obsolete stock declines and warehouse’s capacity is freed up to carry additional products. Consequently, sales forecasting has been the center of attention for academics and practitioners for years. However, there are many hurdles to be taken when it comes to sales forecasting. Basically, the more stable a sales pattern is, the easier the forecasting becomes and the stability of a sales pattern largely depends on the forecast’s level of aggregation, time horizon and stage in a product’s life cycle.

A forecast’s level of entity aggregation determines whether it attempts to forecast sales of multiple products simultaneously or for an individual product and whether it does so for an individual retailer or industry wide. Temporal aggregation refers to the time unit in the forecast, ranging from hourly to yearly forecasts. As a general rule of thumb, forecasting aggregate sales is easier than forecasting disaggregate sales (Dekimpe & Hanssens, 2000). Finally, for individual product forecasts, the product’s stage in its lifecycle poses different challenges. A product typically moves through different stages of sales intensity and stability during its lifetime. Golder and Tellis (Golder & Tellis, 1997) give a useful framework of the different stages in a product life cycle, based on diffusion model theory. First there is commercialization which is the date a new product is sold, consequently there is a period of introduction after which product’s sales (hopefully) take off. A period of growth follows until sales slow down and the product has reached maturity. Finally, sales drop to zero. Although many products might never take off or never reach maturity and the timing of each stage is highly variable, it is a useful definition as it outlines the different sales dynamics that are bound to take place in a product’s life cycle. In general though, forecasting early sales and the height of the initial sales peak is harder than forecasting mature sales. This is mainly due to the unstable sales pattern in early sales and the lack of earlier sales data to build upon. At the same time, early sales forecasting is particularly important, considering that a product’s sales rate may increase by 400% at takeoff (Golder & Tellis, 1997).

**Sales forecasting using traditional online data: web statistics**

The hidden potential in tracking online user behavior was soon acknowledged. As Bucklin & Sismeiro (Bucklin & Sismeiro, 2003) succinctly put it: "Since its commercial inception, one of the most promising aspects of the Internet has been the ability to track the behavior of Web site visitors". Dekimpe & Hanssens (Dekimpe & Hanssens, 2000) recognized the potential of web data in sales forecasting. They envisioned how the availability of web data would
lead to more up-to-date information on changes in consumer interest (e.g. number of website visits), shopping intensity and repeat purchases. The authors stated that this would inevitably lead to better and more specific sales forecasts.

So what is considered traditional web data? Web data is a broad concept that basically encompasses all data that is collected through the organization’s website. These include website statistics (such as page views) and customer input gathered through the company’s interactive elements (such as search queries on the internal search engine). Many academics developed models to predict whether a visitor is going to buy, based on his clicking behavior on the website (Moe & Fader, 2004; Montgomery, Li, Srinivasan, & Liechty, 2004). These studies for the most part use individual choice models to predict individual outcomes. In addition, some studies have been conducted to use this data in aggregate forecasting, but this was usually limited to the prediction of web page traffic itself, so the need for additional computational resources could be predicted (Basu, Mukherjee, & Klivansky, 1996; Li & Moore, 2008).

Sales forecasting using Web 2.0 online data: online product buzz

The importance of web statistics has been accepted by both practitioners and academics. Although not necessarily in sales forecasting, its value is obvious too many practitioners and academics when it comes to evidence-based management and marketing accountability. However, a new trend is emerging which leads to the conclusion that a lot of valuable information is to be found outside the organization’s website.

Whereas the primary function of the Internet has been information retrieval and one-to-one communication (through e-mail) for years, there is a notable shift towards online social networks and public blogs. This “socialization of the Web” makes the volume and intensity of the customer’s opinion much more transparent to anyone with Internet access. The image of the customers no longer needs to be limited to proprietary data collected through an organization’s website. Instead of one-way communication (the company speaks), or two-way communication (the company and customer talk) it is increasingly possible to listen to the conversations potential customers are having with each other.

Never before was it possible to follow a customer’s trail so easily and unobtrusive. As outlined by Dellarocas et al. (Dellarocas, Awad, & Zhang, 2004), customers’ conversations no longer “disappear into thin air”, but are now stored on the Web and publicly accessible. Urban & Hauser (Urban & Hauser, 2004) call this “listening in”. By listening in to the conversations of (potential) customers, there is a lot to learn about their needs and wishes without bothering them with extensive questionnaires. In their research they show how to distill new product opportunities by listening in to ongoing dialogues between customers and Web-based virtual advisers. In addition, the actual “listening in” is facilitated by the fact that many online data aggregation websites such as Technorati and Blogpulse have opened up their data to third parties. Many of them even deployed an API (Application Programming Interface), to ease the extraction of data.

Before the value of this new type of data is further investigated, the following question needs to be answered: are these online expressions representative of actual behavior (either on- or offline)? One might argue that internet savvy consumers preparing for a purchase or action online are likely to be more affected by these online expressions than ‘the average Joe’. However, research indicates that online expressions are indeed a good proxy for actual behavior. Dellarocas et al. (Dellarocas, Awad, & Zhang, 2004) found that average movie ratings online showed high correlation with the ratings of people who rated the movies offline. Godes and Mayzlin (Godes & Mayzlin, 2004) analyzed online newsgroup conversations about television shows and found that the dispersion thereof did indeed resemble actual viewership ratings.

Many expressions are used to describe these emerging data sources: “chatter” (Gruhl, Guha, Kumar, Novak, & Tomkins, 2005), “blogger sentiment” (Mishne & Glance, 2006), “trends” (Glance, Hurst, & Tomokiyo, 2004) or even “online intelligence” (Urban & Hauser, 2004). Dye (Dye, 2000) used the broader expression “buzz” to refer to the phenomenon of a sudden explosion of word-of-mouth promotion of a particular product or story. More specifically, buzz concerns the pace, intensity and reach of word-of-mouth promotion. All these three features have increased dramatically since the internet, explaining the increased emphasis on (electronic) word-of-mouth marketing and the many terms used to describe it.

However, as this study aims to identify the predictive power of different kinds of online data sources and subsequently analyze their predictive value, the above definitions of buzz being almost a synonym for online word-
of-mouth activity, although an important component of buzz, seem too limited. Therefore, this study uses the following definition for online product buzz.

**Online product buzz: any online expression of interest in a product**

This captures both explicit expressions of interest – such as mentioning the product to friends – as interest implied by behavior – such as visiting the product’s web page or looking for it on a search engine.

This online product buzz is interesting in two respects. First, as outlined before, it makes mechanisms that have always been considered to be important sales drivers (such as word-of-mouth) more transparent than ever before. Second, it is not unlikely that potential customers participate in some kind of online product buzz activity, before making the actual purchase. This would imply that buzz measures precede sales, making it a promising predictor. Consequently, in addition to the more traditional sales predictors and web statistics, the following three data sources are taken into account: online reviews, blogs and search trends.

**Online reviews**

Word-of-mouth marketing is generally accepted to play an important role in the success of goods (De Vany & Walls, 1996). As the Web has made engaging in word-of-mouth behavior much easier, word-of-mouth has increased in intensity and visibility due to ever-increasing usage of the Internet (Rangaswamy & Gupta, 2000). Online reviews in particular, have shown to be powerful indicators of the intensity of underlying word-of-mouth (Duan, Gu, & Whinston, 2008).

An increasing body of academic literature exists that show how online reviews are accurate predictors of future sales. After proving how online ratings can reliably be used as a proxy for overall word-of-mouth activity about movies, (Dellarocas & Zhang, 2007) found that the number of reviews posted in the first week proved highly predictive of a movie’s total revenues. This is congruent with an earlier study by Eliasberg & Shugan (Eliasberg & Shugan, 1997) which finds that film reviews have no significant correlation with early box office receipts but do correlate with late and cumulative (total) box office receipts. The two studies seem to suggest that reviews have the most predictive power in a later stage, which makes sense as the product has to be bought before a meaningful review can be written.

Additionally, Duan et al. (2008) find that the volume of reviews proves to be more indicative of future sales than their rating and valence. This in turn is contradictory to research by Chevalier and Mayzlin (2006) on the impact of customer reviews on book sales, which shows that the impact of a negative review on sales is greater than of a positive review, as the most reviews are largely positive.

So, although there seems to be sufficient evidence of the predictive power of online customer reviews, further study is needed. This is even more so, considering that most current research is aimed at “experience goods” such as movies and books instead of actual (retail) products. Reproducing the findings of earlier studies on experience goods for “normal” products would strengthen the belief in the predictive power of online reviews for a variety of products.

**Blogs**

Whereas online reviews are often seen as a proxy for word-of-mouth activity, blogs are often related to product awareness as well. The ‘blogosphere’ can be seen as a representation of the opinion of millions of potential customers (Mishne & Glance, 2006) and is thus likely to be a measure of product awareness and word-of-mouth behavior. Whereas reviews inform those people that are already aware of a certain product, one blog might trigger many readers to consider a product they never heard of before.

Similar to studies concerning the predictive power of reviews, both the volume (Gruhl, Guha, Kumar, Novak, & Tomkins, 2005) and the sentiment (Mishne & Glance, 2006) are shown to have value in sales forecasting.

Sadikov et al. (2009) find that in general, blog references precede movie sales by a week, making it a very interesting sales forecasting predictor. Also similar to online reviews studies, existing literature is found primarily in the movie and book industry, stressing the need for further research in different industries.
Search trends

It is only very recent that analyzing search trends is at all possible. Fueled by Google’s opening up of Google Trends, some very recent studies have been conducted into the predictive power of search engine data, mostly by Google employees. Ginsberg et al. (2008) were able to construct a model that detects influenza epidemics by monitoring influenza related search queries. It is shown that certain queries are highly correlated with the percentage of physician visits. As this data is obtained real time, epidemics are detected with a reporting lag of only one day (in contrast to the current 1-2 week lag of official reports). The validity of search queries as a proxy for actual behavior is confirmed by Choi & Varian (2009) who are able to ‘predict the present sales’ for aggregate sales figures in retail, automotive, housing and travel on the basis of Google Trends search query data.

Research into the applicability of this data source is however in its infancy and the main question that needs to be addressed is whether search queries have any predictive value when it comes to actual forecasting on an individual product level. The two studies mentioned were both only able to report on current activities at an aggregate level.

Methodology

Setting

To examine the predictive power of online buzz in sales forecasting, we used sales data from the Dutch online retailer Coolblue. They currently exploit 50 web shops in the Netherlands and Belgium, each focused on a specific product group. Ranging from Smartphones to shavers, from laptops to drilling machines, they focus on offering a wide range of products and the accompanying accessories. While they conduct most of their business online, they also exploit four physical outlets. It is one of the largest online retailers in the Benelux, with over 80 million euro in revenue (Coolblue, 2009), 195 employees and physical outlets in Rotterdam, Eindhoven, Groningen and Antwerp.

Case selection

Products were selected for the product group ‘smartphones’, as for the studied company this category’s sales exceed all others. The cases were selected on the basis of: data availability (some buzz measures could only be gathered for later time periods), perceived buzz around product launch (the studied organization selected products with highest perceived buzz) and on the basis of total sales (the products with highest numbers were selected, as forecasting is hard for products with only a few sales).

The selected cases include:
- Apple iPhone 3G (introduced on November 1, 2008)
- Blackberry Curve 8900 (introduced on February 26, 2009)
- T-Mobile G2 Touch (introduced on July 16, 2009)
- HTC Hero (introduced on July 1, 2009)
- HTC Snap (introduced on May 7,2009)
- HTC Touch Pro2 (introduced on May 6, 2009)
- Nokia E71 (introduced on July 10, 2008)
- Nokia N95 (introduced on October 29, 2007)
- Samsung Jet S8000 (introduced on June 4, 2009)
- Samsung Omnia II (introduced on July 31, 2009)

Aggregation and time horizon

Sales forecasts come in many varieties. The vast amount of papers written on the topic testifies to its complexity, as each forecasting challenge has unique characteristics. A broad division can however be made on the dimensions of
aggregation and time horizon. As outlined in literature, forecasting aggregate sales is easier than disaggregate sales forecasts (Dekimpe & Hanssens, 2000). This goes for the temporal aggregation (weekly versus yearly) and entity aggregation (product versus product group and industry forecasts versus retailer-specific forecasts). Disaggregate forecasting methods often involve more variables with more complex computations and exhibit a larger degree of instability. If a shop closes unexpectedly for one day, this matters a great deal in the weekly forecast but much less in the monthly or even yearly figures.

Research by Fildes (Fildes, 1989) shows that although individual forecasting methods often lead to more accuracy, they prove to be more complex, time-consuming and less stable as well. The relative impact of predictors in individual forecasts is higher compared to aggregate forecasts, which sometimes causes slight changes in a predictor to cause dramatic effects in the eventual forecast. Consequently, many retailers use an indirect method (Mentzer & Bienstock, 1998): first they forecast aggregate sales for the entire market and then use a calculated multiplier (e.g. market share) to come up with the organization’s specific forecast. In this study, such an indirect method seems inappropriate as the studied organization reports a high variability of market share during the life cycle of a product. Further complicating this approach is the fact that for some products they seem to mainly attract early adopters, resulting in a very high market share in the first weeks after a product launch, while for some more “mainstream” products their market share rises later.

This study assumes that the level of data aggregation and the forecast’s time horizon should be determined by the organization’s business procedures. For the studied organization, purchasing is done on a weekly basis. Using their current forecasting model, the remaining stock in weeks is calculated for each product. When this falls beneath a certain threshold, a purchasing order is made manually. This approach stresses the need for an accurate forecasting model. Consequently, this study generates forecasting models that predict sales for individual products, for an individual retailer, on a weekly basis.

However an analysis of the overall (entity aggregated) sales data for the entire product category ‘smartphones’ was also included for the years 2006, 2007 and 2008 to broaden the applicability of this study’s results to more aggregated forecasting issues. Although this obviously excludes product specific buzz, Google Trends can be downloaded for a collection of search phrases.

**Five test situations across the product life cycle**

As outlined in the literature review, different circumstances pose distinctive challenges to sales forecasting models. For an in-depth analysis it is therefore desirable to test the predictive power of online product buzz in a multitude of test situations. In this study, the predictive power of online product buzz is assessed using multiple models (time series and predictive analytics), for multiple levels of aggregation (product-specific and aggregate) at multiple points in time (from introductory sales to mature sales).

First, the predictive power of online product buzz variables is tested for product-specific sales in each stage of the product life cycle. Loosely based on (Golder & Tellis, 1997), the following stages are distinguished: introductory sales (week of introduction), embryonic sales (week 1 to 4), early sales (week 5 to 20) and mature sales (week 20 to 60). The fifth test situation consists of aggregate sales in the entire product group Smartphones in 2006, 2007 and 2008. This is a useful test case as it represents a more stable sales pattern and it allows a more thorough answer to the overall research question.

As a result, five situations have been constructed:

- Situation A: Product-specific introductory sales: product sales in week 0 (the week of introduction).
- Situation B: Product-specific embryonic sales (from week 1 to week 4 after introduction)
- Situation C: Product-specific early sales (from week 5 to week 20)
- Situation D: Product-specific mature sales (from week 20 to week 60).

Why such a distinction? First, proving the buzz measures’ predictive power for both aggregate (situation E) and product-specific sales (situation A, B, C and D) potentially strengthens the robustness study’s findings. Second, dividing product-specific sales in embryonic (B), early (C) and mature sales (D) allows for more insight into where
the buzz measures add predictive value. It is interesting to see whether online buzz measures are more helpful in overcoming the challenges facing the forecasting of product introductions and early sales, or whether they add predictive accuracy to mature sales, an area in which time series are known to provide fairly accurate predictions. In addition, an attempt is made to forecast the peak sales in the week of a product’s introduction (situation A), which is in fact the purest test of the online product buzz’s value in sales forecasting, since it focuses on an area where there are no historical sales data to base any forecast on.

**Measures**

Based on the literature review, the following variables were collected for the forecasting challenges at hand, collected from company databases, Google and a custom-built screenscraper for the blog/review sites.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description and source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weeks since_stock</td>
<td># weeks since the product was first on stock. Ranges from -4 to + 96</td>
</tr>
<tr>
<td></td>
<td>Source: company</td>
</tr>
<tr>
<td>Sales</td>
<td># sales per week</td>
</tr>
<tr>
<td></td>
<td>Source: company</td>
</tr>
<tr>
<td>Stock Notifications</td>
<td># subscriptions for ‘notify me when on stock’</td>
</tr>
<tr>
<td></td>
<td>Source: company</td>
</tr>
<tr>
<td>CPC Impressions</td>
<td># impressions for Google Adwords campaign</td>
</tr>
<tr>
<td></td>
<td>Source: company / Google</td>
</tr>
<tr>
<td>CPC Clicks</td>
<td># clicks for Google Adwords campaign</td>
</tr>
<tr>
<td></td>
<td>Source: company / Google</td>
</tr>
<tr>
<td>Pageviews</td>
<td># pageviews</td>
</tr>
<tr>
<td></td>
<td>Source: Google Analytics</td>
</tr>
<tr>
<td></td>
<td>Source: Google</td>
</tr>
<tr>
<td>Google Trends Local</td>
<td>Normalized Google Trends data for local search (within Netherlands)</td>
</tr>
<tr>
<td></td>
<td>Source: Google</td>
</tr>
<tr>
<td>Reviews onsite</td>
<td># reviews on organization’s website</td>
</tr>
<tr>
<td></td>
<td>Source: company</td>
</tr>
<tr>
<td>Reviews offsite</td>
<td># reviews on external websites</td>
</tr>
<tr>
<td></td>
<td>Source: vergelijk.nl, kelkoo.nl, elcheapo.nl, tweakers.net (four major Dutch review and comparison shopping websites)</td>
</tr>
<tr>
<td>Blogpulse</td>
<td># new blogposts per week</td>
</tr>
<tr>
<td></td>
<td>Source: Blogpulse</td>
</tr>
</tbody>
</table>

An important distinction can be made here between "internal" measures that could be collected relatively easily as they were owned by the company and accessible through their own system (such as previous sales, stock notifications, onsite reviews and, through an API, the Google Adwords and Google Analytics). Collecting "external" data such as Google Trends data, offsite reviews and Blogpulse data proved much more time consuming and complex. These data sources were also more error prone as more data ‘cleansing’ was needed.
Model development

A common characteristic of predictive analytics is the almost religious use of holdout samples. Predictive analytics is often used when there is plenty of data available. Typically, the model is trained on a large dataset with known response values (the training set), which is also known as ‘supervised learning’ (unsupervised learning models are not currently of interest). Subsequently, the models are improved using a validation set and their predictive accuracy is measured on a holdout sample. In practice, all available models are developed and a choice is made based on the reported accuracy of their predictions (Shmueli, Patel, & Bruce, 2007). This means that in situation A, B and C (or: early sales) the model was estimated for all products but the one in the holdout sample and subsequently cross-validated. In situation D and E, the model was estimated on prior sales data, and future sales were held out.

For practical applicability, a predictive technique is selected that (a) outputs the relative importance of each variable, and (b) is intuitively understood by the decision maker. The most obvious candidate would be a Regression Tree as this technique satisfies both criteria. Regression trees distill “rules” on the basis of training data and automatically extract the best ‘split points’. This is done using the Gini impurity criterion (Shmueli, Patel, & Bruce, 2007). A typical rule could be: if sales last week where above 20 and the number of new blog posts is higher than 10, sales will be 80. These regression trees are built using recursive partitioning of the underlying data. Each split point in the tree is based on the extent to which they reduce the impurity (or: heterogeneity) in the descendent nodes (or: leafs). For regression trees, the typical impurity measure is the sum of the squared deviations from the mean of the leaf (Shmueli, Patel, & Bruce, 2007).

However, as we study a relatively large number of variables and have only limited sales data, building a single tree proves unreliable and greatly dependent on the underlying training sample that is randomly selected. To counter this common problem, Breiman (2001) devised the method of random forests. This method, much like cross-validation, generates a lot of regression trees and each tree is constructed using a different bootstrap sample from the original data and tested on the left-out sample (the ‘out-of-bag error estimate’). For each variable used, an importance score is constructed using the number of times a ‘split’ was made on this particular variable. After building 1000 regression trees, output gives the average importance of each variable (Breiman, 2001). The importance of every variable is calculated as the extent to which every variable decreases the impurity – calculated as outlined before – of its subsequent nodes. This number is added up for all the trees generated and outputs a number which is known as the Gini importance measure (Breiman, 2001). As such, a ranking of the importance of each variable is given. Note that although the absolute values of the Gini importance measures have no meaning, and thus the model loses some interpretability compared to a standard regression tree, the ranking of the variables’ importance still gives valuable information as to which type of data is more or less useful for prediction. More importantly though, just as cross-validation is found to be a good way of overcoming data limits in more traditional forecasting models (Cooil, Winer, & Rados, 1987), random forests clearly outperform individual regression trees when data is limited (Breiman, 2001).

Model assessment

Models were assessed using the Root Mean Squared Error (RMSE) measure. This accuracy measure is common in forecasting literature and its calculation is shown in the following formula.

\[ \text{RMSE} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{(A_i - F_i)^2} \]

A denotes actual sales, F the forecasted value and n the number of observations.

The comparison of models was made on the base of the increase (or decrease) in accuracy between two models is calculated as the percentage change in RMSE. This was calculated as follows:

\[ \Delta \text{Accuracy} = \frac{1}{n} \sum_{i=1}^{n} \frac{\text{RMSE}_A - \text{RMSE}_B}{\text{RMSE}_B} \]

The random forest models with online product buzz variables were compared to the random forest models without these variables. Additionally, these models were compared against the company’s baseline forecasting model, which is a simple weighted moving average model, as outlined below. Additionally, all developed models are compared
against the organization’s baseline model, which is in essence a simple weighted moving average (Axsater, 2006). Note that the coefficients have been left out due to the sensitivity of this information.

\[ S_t = \varphi_1 S_{t-1} + \varphi_2 S_{t-2} + \varphi_3 S_{t-3} \]  

(1)

**Results**

Without exception, it was found that including the online product buzz variables in the random forest model improved the model’s forecasting accuracy. On average, the forecasting error decreased by 27.78%. The relative improvement is particularly high in very early sales forecasting situations (introductory and embryonic) where accuracy is improved by about 42%.

Additionally it was found that the random forest with the buzz variables also outperformed the company’s baseline model by 30.23%. Without buzz variables, the random forest model still outperforms the baseline model in introductory to early sales (situations A-C), but not in the more stable situation (D and E).

**Variable importance**

The assessment of the relative importance of each variable is solely based on the output of the random forests. As the absolute Gini importance values have no meaning, the importance measures in the table were normalized so that the most important variable has a value of 100. All variables above 50 have been made bold to facilitate interpretation. Additionally, the relative ranking has been included between brackets. Note that S(t-1), S(t-2) and S(t-3) denote the actual sales from respectively 1, 2 or 3 weeks ago (or pre-orders if they were placed prior to the launch date).

<table>
<thead>
<tr>
<th>Table 2. Variable importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>S_{t-1}</td>
</tr>
<tr>
<td>S_{t-2}</td>
</tr>
<tr>
<td>S_{t-3}</td>
</tr>
<tr>
<td>StockNotifications</td>
</tr>
<tr>
<td>OnsiteReviews</td>
</tr>
<tr>
<td>OffsiteReviews</td>
</tr>
<tr>
<td>GoogleTrendsLocal</td>
</tr>
<tr>
<td>GoogleTrendsGlobal</td>
</tr>
<tr>
<td>Blogpulse</td>
</tr>
<tr>
<td>Pageviews</td>
</tr>
<tr>
<td>CPCClicks</td>
</tr>
<tr>
<td>CPCImpression</td>
</tr>
</tbody>
</table>

Some noteworthy results are found in table 2. First, the predictive power of prior sales data is only really recognized in early, mature or aggregate sales. Until week 5 in the product life cycle, it has hardly any predictive power. Second, StockNotifications and GoogleTrendsLocal are very important in introductory and embryonic sales, but lose practically all predictive power in later stages. Third, OnsiteReviews and OffsiteReviews show the same pattern as prior sales data, illustrating high predictive power in situation C, D and E, while have almost none in situation A and B. Finally, Pageviews is a relatively consistent, well-performing, predictor.
Discussion

The basic empirical result is that the predictive power of online product buzz variables is clear (average accuracy improvement of 28%). However, the results for the different phases of the product life cycle also have theoretical implications for what drives or precedes sales in each phase.

Variable importance: first watch the innovators, then the imitators

Bass (1969) based his adoption model on the concept of innovators and imitators. First, he stated, innovators or ‘early adopters’ would adopt a new technology or product out of curiosity. Only after these innovators have some experience with the product, the word will spread and the imitators will start buying. In a recent study, Young (Young, 2009) attributed this phenomenon to the concept of contagion, social influence and social learning.

It requires little imagination to recognize this innovation-imitation pattern in the relative importance of the online product buzz variables in this study. In the very early stages, only variables that indicate individual orientation behavior – typically shown by innovators – demonstrate notable predictive power: subscriptions to stock notifications, pageviews and local search trends. However, when the product moves towards its 5th week of sales a shift happens and on- and offsite reviews – indicative of word-of-mouth activity (Duan, Gu, & Whinston, 2008) and thus imitation – start to exhibit high predictive power. Or, relating to the concept of online product buzz, in early sales the implicit interest shown by behavior is more predictive. In later stages the explicit expressions of interest become more important.

And although predictive accuracy has little to do with whether the model is true (Hitchcock & Sober, 2004), demonstrating that an already existing theory also has predictive validity is often found to reinforce its credibility (Shmueli & Koppius, 2010).

Prior sales (or: lagged sales variables)

Over the years, the extrapolation of prior sales numbers into the future has produced remarkably accurate results (Axsäter, 2006). As is stated by many academics, this is particularly so for stable sales patterns, and thus mature or aggregated sales (Hanke & Wichern, 2009). This claim is supported by this study’s finding in that the lagged sales variables are of limited relative importance in introduction and embryonic sales (ranked 4th), while being the most important in situation C, D and E. This emphasizes the potential value of capturing buzz variables for very early sales forecasting.

Preorders and Stock Notifications

As found by Hui et al. (2008) and Moe & Fader (2002), preorders can have high predictive power in sales forecasting. Preorders, although limited, were included in this study as regular sales in the weeks before the product’s introduction (i.e. when it came on stock). However, at the studied organization visitors also have the opportunity to receive an e-mail when the product comes on stock, called Stock Notifications. An analysis of the relative importance of each variable in situation A, B and C shows that these Stock Notifications prove one of the most important predictors for introduction, embryonic and early sales and have more predictive power than preorders. The explanation for this is probably twofold. First, the barrier to subscribe for a stock notification is presumably lower than for an actual preorder. This makes it an interesting and valuable variable to capture customers’ interest. Second, because of this lower barrier, more stock notifications have been captured for the studied products, making it a more exploitable predictor than preorders. As should be expected, the predictive importance of Stock Notifications drops to 0 in later stages.

Pageviews

Another important variable to consider in early sales forecasting is the number of last week’s pageviews. It seems likely that the number of pageviews is a reliable proxy for customers’ interest and that people visiting the product page eventually decide to buy. However, due to the absence of other reliable predictors such as prior sales, this variable is in fact the most important to take into account in very early sales (situation A and B). The potential hereof is even higher, considering that any online retailer already captures this variable through regular web statistics. In situation C and D its relative importance decreases considerably. Surprisingly, hardly any studies were found that incorporate pageview statistics in sales forecasting. Clickstream measures such as these have previously
only been used in individual choice models (e.g. Bucklin & Sismeiro, 2003 and Montgomery A., Li, Srinivasan, & Liechty, 2004) and the forecasting of pageviews itself to account for sudden changes in required networking capacity (e.g. Basu, Mukherjee, & Klivansky, 1996).

**CPC Clicks and Impressions**

The relative importance of Google Adwords impressions and clicks is stable for all situations. Of all 12 predictors, CPC Clicks consistently ranks 6th in terms of Gini importance and CPC Impressions last or 11th. Although this does not imply anything about the value of CPC campaigns, it does downplay its relative importance in the forecasting process.

**On- and Offsite Reviews**

Many academics demonstrated the predictive power of online reviews (i.e. Dellarocas, Awad, & Zhang, 2004; Duan, Gu, & Whinston, 2008; Eliashberg & Shugan, 1997). In this study it was shown that the relative importance of reviews is very high, particularly in situation C and D (respectively ranked 4th and 1st). This study thus confirms the relative importance of on- and offsite reviews in sales forecasting and contributes two important findings.

First, the relative importance of online reviews is highest in later sales. This makes sense, considering that buyers of the product need to have some experience with the product before they are able to write a review. It also confirms findings by Eliashberg & Shugan (1997) in “experience goods”, that reviews have no significant correlation with early sales (in their case: box office receipts), but do correlate with late sales.

Second, it is found that onsite reviews prove more predictive than offsite reviews. A plausible explanation for this is the fact that the number of people writing reviews on the organization’s website is an indicator of the amount of people that did already buy it from the studied organization. This is interesting, as capturing the number of onsite reviews is far less complex than capturing offsite reviews.

**Google Trends (Local & Global)**

This study shows the value of Google Trends search data in sales forecasting. Two interesting conclusions can be drawn. First, the predictive power of search trends is highest in introduction and embryonic sales (both ranked 3rd). This is partly due to the absence of other predictors, but might also be explained by the fact that when a new product is introduced an increasing trend in search behavior is found. The second conclusion is essentially a confirmation of what was logically expected: local search trends prove more predictive than global search trends. As local searchers will be the eventual buyers, it was expected – yet not confirmed before – that local search trends would be a better indicator of future sales.

However, Choi & Varian’s (2009) findings on the predictive power of Google Trends data in aggregate sales forecasts were neither confirmed in one-step-ahead forecasts nor in ‘predicting the present’. This might be due to a misfit between the captured search behavior (all phrases related to ‘mobile phones’) and the studied objects (‘smartphones’) or simply be a disconfirmation of their claim. Future research is needed on this matter.

**Blogs**

Earlier studies by Gruhl et al. (2005) and Mishne & Glance (2006) find that blog posts have considerable predictive power in sales forecasting. The results of this study do not refute this, but do temper expectations of its added value. Although only blog post volume was recorded – leaving out the sentiment of those posts – the relative importance of blog data in sales forecasting seems limited, judging it’s 7 out of 10 ranking of importance. A possible explanation might come from the fact that both prior studies were focused on books. As books have a longer product life cycle, the ‘stream of blogposts’ might be more stable and rich than for smartphones. A more abstract explanation – that cannot be checked in this study – might be that bloggers have more authority when it comes to books than when it comes to smartphones. Third, an important note is that the blogs studied were mostly from the U.S. with little coverage of Dutch blogs that might possibly be more influential, making a strong correlation with sales in the Netherlands less likely. Finally, it could be that the earlier studies had a different definition of predictive power. As both only checked for a correlation between blog posts and actual sales, it might be that for their datasets an out-of-sample forecast would also exhibit poor performance.
Managerial relevance

The business value of monitoring online product buzz

The business value of measuring the online product buzz variables is interesting to get some perspective on the impact of these accuracy improvements. A modest approximation of the business value can be made as follows. In situation E (aggregate Smartphone sales), it is shown that the RMSE decreases from 171.66 to 157.68. Overly simplified, this means that on weekly basis 14 products more can be sold – because they were formerly not on stock – or 14 less products have to be kept on stock. Per product we assume $200 dollar in extra revenue. In consumer electronics, the risk of obsolescence is high and consequently the costs of inventory are assumed at 20% of the purchasing price, per week. For the studied organization and the studied product group, the business value of this increased accuracy thus ranges between $560 and $2,800 a week, or $29,120 - $145,600 on yearly basis. Per product, a simplified approximation of the business value can be calculated as follows. The difference in RMSE for each situation is multiplied by the number of weeks in that situation and the purchasing price ($200) for the upper limit and multiplied with the costs of inventory ($40) for the lower limit. Subsequently, the business value of all situations is summed up. Following this calculation, the business value of using online buzz measures in sales forecasting ranges between $3,789 and $18,944 per product.

Potential solution to early sales forecasting difficulties

As Aaker & Jacobson (1987) noted – in their comment on Brodie & De Kluyver (1987) – part of the strength of naïve models is that historical sales data inhibits implicit information on a product’s popularity and adoption. Or as Golder & Tellis (1997) put it, once naïve models start to perform it is already known if ‘it will ever fly’. And even better: it is by then already known how high it will fly. Due to the absence of such information, in early sales forecasting one is sentenced to look for other cues about a product’s future sales. Online product buzz variables are, as expected, found to inhibit such predictive power. As shown earlier, the improvement in accuracy was highest in introductory and embryonic sales. This makes sense, as time series models are by then still highly unstable and inaccurate, due to inadequate data. Just as Moe & Fader (2002) showed the predictive power of preorders, this study clearly shows that online product buzz variables are a potential way to overcome the difficulties of (very) early sales forecasting. Additionally, it targets further research into this topic in the direction of more ‘goal-oriented’ (innovation) measures, as ‘imitation’ measures only render useful in later stages of the product life cycle.

Conclusion

The implications of this study are of practical, methodological and theoretical nature. Theoretically, it is shown that expected drivers of sales such as explicit word-of-mouth activity (captured in online reviews) and orientation behavior (captured by search trends and pageviews) also have predictive validity. Somewhat surprisingly, blog data did not demonstrate predictive power (although some possible explanations were offered earlier in the paper), whereas prior research did establish its predictive power.

More interestingly however, is an assessment of the predictive value of each of the predictors in different stages of the sales cycle. It is shown that ‘innovation’ behavior is much more predictive in early sales, whereas ‘imitation' behavior is of bigger value in later stages. This is interesting as it strengthens Bass’s (1965) concept of innovation and imitation, and therefore suggests a longitudinal effect of ‘buzz’. It is gives a more elaborate view of the different types of online product buzz that should be studied for different purposes.-

Methodologically it was found that a random forests technique is very suitable to incorporate these online product buzz variables. First, the nonlinearity of random forest models allows for complex relationships between independent and dependent predictors. Second, the data that is gathered from the web is often of enormous quantities and displays a lot of missing values and occasional outliers. The tolerance of random forests makes forecasting a lot less painful.

Practically it is shown that there is real evidence of the value of online product buzz in sales forecasting. Especially in early sales forecasting, the potential upside of including these variables in the forecast are big. However, it is also
found that monitoring these variables is rather labor-intensive, thus prudence is advised in the number of data sources that are captured. This study suggests that one is best off monitoring stock notifications and pageviews and local search trends for early sales, while capturing onsite reviews for mature sales.

**Focusing on the low hanging fruit**

Obtaining online product buzz data is not without costs. Particularly public data requires effort to incorporate. The Google Trends data for example, was gathered by manually downloading CSV files with the trend data for each product, for each month. Other public data, such as offsite reviews and blog data were either collected through building a ‘screen scraper’ or through communication with an API. Both approaches take effort to develop and can suddenly be rendered useless, when the public data aggregator decides to change its website or API specifications. Therefore it is advised to start out by only gathering the data with highest predictive power and the lowest effort required to obtain. For early sales, a lot of predictive power can be added by (a) extracting pageview statistics from web analytics software and (b) offering the opportunity to subscribe for stock notifications. Both of these data sources are internally managed, which secures their reliability. As the costs – and thus the potential gains – of wrongly forecasting early sales are relatively high, it might be worthwhile to capture search trends data as well. However, this would require a screen scraper that automatically downloads a CSV file and rightly processes the content in it.

For later sales, a great improvement is expected when onsite reviews are collected and incorporated in model development. This obviously means that if from a sales forecasting perspective, it is advised to include the opportunity of posting reviews on your website. Additionally, utilizing pageview statistics again seems worthwhile and easy. The same trade-off needs to be made here regarding offsite reviews. Is it worth collecting this against the expense of developing (and maintaining) the necessary screen scraper? The decision of whether the efforts of actively harvesting these external online product buzz measures depends on an organization’s volume – and thus the potential upside – and its technical capabilities in-house. In the case of the studied organization, both criteria suggest it might be worthwhile to capture external sources as well.

**Start forecasting with a robust forecasting technique**

When one tries to get his feet wet in sales forecasting, it is easy to drown in the variety of models, statistical validity measures, and other concepts such as ‘detrending’ and ‘seasonality’. Although all of these concepts are important, it all too often hinders any attempt to start forecasting. This study shows that random forests exhibit some very interesting features that could dramatically lower the barrier to forecasting situations with enormous amounts of data - such as in online retail. Its tolerance of ‘noisy’ data input, its prime focus on predictive accuracy (instead of building the "truest" model) and the relatively short time it takes to get up to speed make it an interesting candidate, especially for more ‘unstable’ forecasting situations as early sales. Also, an increasing amount of studies displays that its predictive accuracy equals or exceed those of more traditional forecasting models.

**Limitations and further research**

The first and most prominent limitation of this study into online product buzz is the absence of two of its most notable proponents: Facebook and Twitter. Twitter was not included because at the time of data collection, the limit of searchable updates is around 1.5 weeks (Twitter, 2010). As data collection was started after the actual sales period took place (a posteriori) it was impossible to obtain the necessary data. Requests made to Twitter received no response. In future research, it is advised to start collecting Twitter data early by harvesting the data on a weekly basis for the studied period. Facebook was left out of the research because at the time of data collection, no valuable data was extracted from its search functionality. Mainly due to Facebook’s privacy limitations (which have been changed at the time of writing), only very few status updates were searchable. For example, the search phrase “HTC Hero” resulted in 2 hits over its entire product life cycle. As privacy settings have been changed since – status updates are now by default searchable – the value of Facebook “buzz” is expected to increase. Further research might involve Facebook as a data source.

Besides this limitation, there is a need for further research into the value of blog data as well. Not only was this study unable to confirm earlier reports of improved predictive accuracy by measuring blog activity (Gruhl, Guha, Kumar, Novak, & Tomkins, 2005) and neither has it tested the claim that measuring sentiment improves forecasts (Mishne & Glance, 2006). More specifically, the potential value of blog posts in predicting the height of a product’s peak sales at introduction was not examined due to missing data. In general, this research would benefit from more
data. Following Hanke & Wichern (2009), mainly for situation A, stronger conclusions could be drawn if the dataset would contain at least 130 records.

Finally, a potential selection bias occurred in the case selection. As products were selected on, among others, their perceived ‘buzz’, the results of this study might overestimate the predictive power online product buzz in sales forecasting. However, the dataset showed that for three products (the T-Mobile G2 Touch, Samsung Jet S8000 and the Samsung Omnia II) very little buzz was collected. Consequently, rather by luck than effort, the selection bias seems limited and the results are not expected to present an overly optimistic view of the findings of this study.

The specific goal of this study was to assess the predictive power of online product buzz. As a result, no attempts are made to identify the underlying causality. Many questions remain standing: do blog posts drive people to buy or is it merely an indicator of how many people bought already or an indicator of for instance an effective advertising campaign by the seller? Are people more influenced by onsite or offsite reviews and why? What causes “buzz” and for which product is it more likely to occur?

Further research is needed to study these questions to more precisely identify meditational effects and multicollinearity issues that can aid in further theorizing. At the same time, for practical purposes it is irrelevant if the reviews in week 1 actually cause sales in week 3 or only predict sales. This is a critical but often overlooked distinction between developing a model and developing a predictive model, where including ‘non-causal’ variables in a model can sometimes lead to better models than ‘causal’ variables (Shmueli & Koppius, 2010). This does not take away the notion that building a purely explanatory model to gain more insight into the underlying causal mechanisms of retail sales and the role of online product buzz would make an interesting avenue for future research.

A second theoretical question that remains open is to what extent similar products (in the same ‘product group’) demonstrate similar sales patterns. Or more abstract: what makes products similar? This is relevant, as in situation A, B and C the forecasting model is estimated on the basis of 9 products, and used to predict sales for the 10th product. Although this proved to be an effective workaround for the lack of historical sales data in these situations, forecasting models would benefit from a clearer view on what products exhibit similar sales patterns and, more importantly, how one is to know. Conjoint analysis of the features of all products within a product group might find that certain characteristics are related to similar sales patterns.

If it is possible to make an estimated guess of what earlier displayed sales pattern the newly introduced product is expected to resemble, this would probably increase prediction accuracy. Although Fildes (1989) showed that selecting the proper sales model a priori was hard (if not impossible), Sood & Tellis (2009) recently seemed to do just that with their approach of gathering and comparing typical sales curves. Taking the sales curves of product introductions as an object of further research is likely to improve our performance in very early forecasting, thus complementing the role of online product buzz.

References


