Smart Data Works

Delivering ideas for product improvement and shopper understanding

Nikolai Reynolds • Stefan Maas • Robert Wiggall

Copyright

About ESOMAR

ESOMAR is the global voice of the data research and insights community, representing a network of 35,000 data professionals.

With more than 4,900 members from over 130 countries, ESOMAR's aim is to promote the value of market and opinion research in illuminating real issues and bringing about effective decision-making.

To facilitate this ongoing dialogue, ESOMAR creates and manages a comprehensive programme of industry specific and thematic events, publications and communications, as well as actively advocating self-regulation and the worldwide code of practice.

ESOMAR was founded in 1948.

About ESOMAR Membership

ESOMAR is open to everyone, all over the world, who believes that high quality research improves the way businesses make decisions. Our members are active in a wide range of industries and come from a variety of professional backgrounds, including research, marketing, advertising and media.

Membership benefits include the right to be listed in the ESOMAR Directories of Research Organisations and to use the ESOMAR Membership mark, plus access to a range of publications (either free of charge or with discount) and registration to all standard events, including the Annual Congress, at preferential Members' rates.

Members have the opportunity to attend and speak at conferences or take part in workshops. At all events the emphasis is on exchanging ideas, learning about latest developments and best practice and networking with other professionals in marketing, advertising and research. CONGRESS is our flagship event, attracting over 1,000 people, with a full programme of original papers and keynote speakers, plus a highly successful trade exhibition. Full details on latest membership are available online at www.esomar.org.

Contact us

ESOMAR

ESOMAR Office:
Atlas Arena, Azië Gebouw
Hoogoorddreef 5
1101 BA Amsterdam
The Netherlands
Tel.: +31 20 589 7800

Email: customerservice@esomar.org
Website: www.esomar.org

Smart Data Works

Delivering ideas for product improvement and shopper understanding

Nikolai Reynolds • Stefan Maas • Robert Wiggall

Introduction

In an increasingly digitalized world, significant levels of potentially valuable data are available with more continuously being produced all the time. However, with this opportunity there is a significant challenge to manage it and to differentiate between smart relevant and non-relevant data. The research industry provides a good example where one of the goals is to identify relevant and valuable insights in consumer and product research using this smart data. Our goal was to conduct a study on a research project using 'big data' and to compare the outcome of the analyses with traditional survey research. Online shopper ratings and review data in social media is an exciting and 'on trend' data source and was compared to traditional survey data. The survey data included product tests, i.e. products were placed in-home and consumers evaluated the products using a standardized questionnaire. The objective was to derive substantive insights about the core drivers for a five-star-rating of consumer reviews in online shops or platform ratings for pet care products and compare these insights with drivers of liking from traditional research on pet care products already existing in the market. We validated the hierarchy of drivers for the overall product rating by conducting a meta analyses on previous product tests and assessed the drivers of overall liking.

Background

Market research has always been aimed at helping consumers and businesses to fully understand facets and deductions of current and future needs and behavior. In today's world, market research has the possibility to build these insights on different types of data which are to some extent available in abundance. However, more data does not mean better data. The nature of the available data is diverse and its validity, use and resulting analyses need to be carefully considered.

In social media and e-commerce there are some special challenges. Some consumers only select and purchase a product online if there are at least a certain number of reviews and the product rating has a certain positive threshold in evaluations. Other consumers may be more skeptical and find it dubious when a product exhibits many positive product ratings or reviews because they suspect that there has been manipulation. In addition, to what extent the reviews are considered in the product choice decision may depend strongly on the category and the knowledge of the consumer about it. Another driver for considering the reviews is of course the price of the product and the number of products to be purchased and the amount of time a consumer has available to purchase a specific product.

Against this background of social media data and ratings and reviews in e-commerce, many established procedures in market research and analytics need to be reconsidered and applied in a different way as the new and sometimes very large amounts of data reflect and influence consumer behavior in a different way than data from ordinary traditional questionnaire research. In addition, there are challenging requirements not only to the software but also to the hardware to be considered when thinking about conducting an analysis on mass data or when combining different types of data.

The need for efficient technology driven approaches on mass amount of data can be facilitated by employing artificial intelligence and behavioral consumer science. The relevance of big data, behavioral, artificial intelligence and technology driven approaches is exhibited in a recent GRIT report where the industry was asked what areas of research provide further opportunities.

Figure 1. Opportunities within research
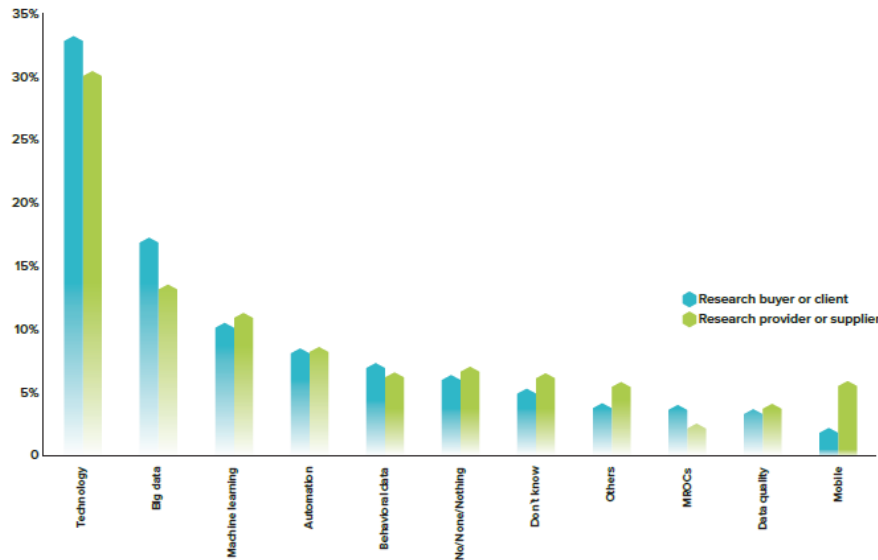
Source: GRIT Report, 2017

Research approach

In this study we compared ratings and reviews from e-commerce platforms to a meta analyses, conducted on past traditional product tests. The past traditional product tests were so called post launch benchmarking projects where existing in-market products are benchmarked against competitor projects. Consumers tested monadically two products: a client product versus competitor product. To be able to compare social media data with survey data, the handling and analyzing a mass amount of data required a combination of artificial and human intelligence.

Product ratings allow a proximate investigation of consumer language. The product ratings form not only a different data source from traditional survey data but at the same time offer an unprecedented view to the consumers and their needs. For product managers they are a valuable source of the level of product acceptance and disclose potential areas for improvement of one's own product (in direct competition comparison). The data furthermore may inspire innovation and communication content for advertisement. They are often so valuable they are often being used directly as a reference in the product development department.

Assorted solutions from Ipsos enable product and service providers to better understand the various marketplaces and thus develop strategies for effective market development. This is achieved by building upon real consumer behavior and their dynamic needs. In addition to the traditional ways of primary research, product intelligence (PI) analytics were applied, i.e. the survey and analysis of product reviews and ratings from e-commerce platforms.

The basis for each PI project is a sufficient number of usable Product Reviews. The number and nature of the reviews determines the analytical breadth and depth of the approach. In some cases, a qualitative content analysis of 50 reviews is adequate, however, often the number of cases is significantly higher, and the evaluation is done through a Machine Based Learning approach. The data collection (Scraping) as well as cross-shop mergers (Mapping) form the basis of the analysis. Data collected, other than the star ratings and review content, is individually tailored to the research objective and client needs. Frequently collected meta-information are date, author name and purchase verification. The latter is considered in the context of the "Fake-Checks", i.e. a review to what extent a genuine review is available and can be used for the analyses. The author names were then deleted, and the data anonymized.

After the data collection (scraping) and mapping the analyses is conducted. The analyses usually include two perspectives.

1. Basic analyses of descriptive data such as distribution, averages or Top2 / Bottom2 boxes of star ratings as well as their longitudinal development over time are investigated.
2. Content of the review texts. Here, we consider qualitative content analysis procedures as well as procedures in the field of Machine Based Learning, such as for example, text mining or modern driver analysis.

If the number of available review permits it, we maximize knowledge gain and insights, by combining these two methods with each other. Our experience has shown that the mechanical approaches behind the analyses help us to identify and map the relevant to product properties using the respective sentiment and allow to understand the drivers for the star rating precisely. In advance to the analyses a corresponding training of the machine must be considered of course.

A comparison of human and machine coding showed that by means of artificial intelligence a similarly high quality can be achieved much faster and therefore cheaper. However, qualitative research provides contextual and transversal insights to complement the quantitative analyses results and derive strategic recommendations for action from the analyzes.

In collaboration with Mars Petcare we analyzed review and ratings from assorted online shops. The shops considered in the analysis constitute a mix of general retailers and smaller, more specialized pet food retailers. All shops were identified via desk research and aligned with Mars and selected arbitrarily depending on the sampling requirements. All shops considered have been online for a comparable amount of time: between 16 and 26 years. ASDA showed exceptionally high review volumes between 2015 and 2017 which is due to an automatic integration of reviews gathered from different brand websites such as whiskas.co.uk. Therefore, the data collection encompassed "crawling" of all available reviews to cat and dog food products from the five UK online shops considering a time horizon ranging from 2015 to 2017. Finally, we considered a data set of n=20,230 reviews.

The descriptive analyses considered volumes per year, shop, food types, packaging technology and considered the Top2 and Bottom2 assessment of brands and products. A text mining system was manually trained by researchers to automatically explore product features according to the Mars Key Performance framework for product performance but considered additional identified topics based on all review texts by using text mining (Natural Language Processing) and qualitative procedures (Grounded Theory). A random forest regression analyses was applied to understand the ranking of the drivers of liking. Qualitative was performed on cases where data volume was too low for accurate quantitative assessment. The synthesis and reporting comprised results from all previous research steps were joined and matched with each other.

Figure 2. Considered e-commerce platforms and reviews

Source: Own compilation

Finally, we compared the research results of the analyses of ratings and reviews to a meta analyses on past product tests on pet food. The meta analyses of past survey data comprised 19 studies encompassing each a set of products that were tested across dog and cat food. A random forest regression analyses was conducted on the survey data to disclose the ranking of drivers of liking.

Table 1. Overview of pet care product testing studies considered

Pet	Food Type	Brands
Cat	Wet (30 products)	Felix, Gourmet, Hill's, Royal Canin, Sheba, Whiskas
Cat	Dry (16 products)	Friskies, Go Cat, Hill's, Kitekat, Purina, Royal Canin, Whiskas
Cat	Care & Treats (12 products)	Felix, Whiskas
Dog	Wet (14 products)	Alpo, Beneful, Cesar, Luvsome, Pedigree
Dog	Dry (4 products)	Nutro, Pedigree, Purina, Blue Buffalo
Dog	Care & Treats (2 products)	Pedigree, Milk-Bone

Source: Own compilation

Findings and Conclusion

The e-commerce data enabled a better understanding of the consumers' language in product and brand perception and supported recommendations for product improvement. For achieving this, we have leveraged and combined both human capabilities and AI (NLP, Machine Learning). Based on this case study conducted together with Mars Petcare we showed the value of the approach on strategic as well as operational level. In addition, the ranking of drivers of liking compared favorably to the meta analyses of survey data on pet food.

In summary the benefits of the approach were:
- High data volumes relating to immediate consumer feedback
- Ability to track trends and reaction overtime
- Supports interpretation of star ratings with commentary e.g. does 5 stars mean 'perfect'??
- Identified the important of clean data over big data….
- Get close to the language that consumers use and interrogate it in a smarter way
- Granular interpretation of what people say to get a true reflection on sentiment towards your product offering

Potential areas of application:
- Support product performance evaluation and monitoring
- Tracking tool to assess trends over time
- Using familiar consumer language when talking about the products online…breaking down barriers to understanding
- Explore different aspects of a specific topic. E.g. What does enjoyment mean and what is it linked to?
- Review your new product development as well as competitors to identify performance drivers
- Get greater understanding of those small niche brands that are normally harder to research

The Authors

Nikolai Reynolds is Head of Global Product Testing, Ipsos Innovation, Global.

Stefan Maas is Practice Lead Social Intelligence Analytics, Ipsos Germany

Robert Wiggall is Global Product Performance Manager, Mars Petcare, United Kingdom