

Bilgram, Volker; Gluth, Oliver; Piller, Frank T.

Article

Crowdfunding Data as a Source of Innovation

Marketing Review St.Gallen

Provided in Cooperation with:

Universität St. Gallen, Institut für Marketing und Customer Insight

Suggested Citation: Bilgram, Volker; Gluth, Oliver; Piller, Frank T. (2017) : Crowdfunding Data as a Source of Innovation, Marketing Review St.Gallen, ISSN 1865-7516, Thexis Verlag, St.Gallen, Vol. 34, Iss. 3, pp. 10-18

This Version is available at:

<https://hdl.handle.net/10419/275909>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Marketing Review St. Gallen



3 | 2017 **SCHWERPUNKT** Interview mit Jochen Werne, Bankhaus August Lenz & Co. AG • Crowdfunding Data as a Source of Innovation • Sketching the Future of Mobility with Visualizations • Customer Insights in Social enterprises • Business Transformation: Customer Insights und unternehmensinterne Denkmuster • Customer Insights in Fintech-Unternehmen • Agiles Marketing digitaler Produkte • Studying Shelf Framing With Multiple Consumer Insights **SPEKTRUM** Social Selling: Status Quo

www.marketing-review.ch

Customer Insight Driven Business Transformation

**BUSINESS TO
CUSTOMER**

Crowdfunding Data as a Source of Innovation

In the digital era, crowdfunding platforms serve as a litmus test for the desirability and viability of thousands of innovative ideas. Based on consumers' actual purchasing behavior, crowdfunding data promises to reveal future hot topics and technologies. This study shows how companies can identify innovation opportunities from crowdfunding data by applying machine learning techniques.

Dr. Volker Bilgram, Oliver Gluth, Prof. Dr. Frank Piller

Open innovation approaches leveraging the creativity and knowledge of external actors such as consumers or more innovative individuals such as lead users have become a prevalent modus innovandi for firms to create value (Boudreau/Lakhani 2013). Companies embrace open innovation in various phases of the innovation process ranging from creating early stage ideas (Brunswick et al. 2016) to developing sophisticated solutions (Jeppesen/Lakhani 2010). While there is an abundance of open innovation methods and tools to actively engage individuals from outside the company in idea generation, concept testing as well as problem solving, approaches for opportunity identification and need exploration in the early phases of innovation endeavors are rather underrepresented. At the same time, we live in an age of big sensing, meaning that the amount of consumer data (e.g. user-generated content on social media or sensor data from mobile devices) is constantly increasing at an enormous speed just waiting to be capitalized on. Despite the enormous potential of these data, companies are missing out on significant innovation opportunities (Roberts/Piller 2016). The negligence of data- and consumer-driven opportunity identification in the open innovation domain is particularly surprising as defining the innovation space and exploring innovation opportunities appear to be tasks best addressed with a clear focus on the consumer in his/her role as the primary need owner (Bilgram et al. 2013).

Opportunity Identification – Current Methods and Their Downsides

In order to ensure customer-centric innovation, market research has long employed methods to integrate the voice of the customer through surveys, focus groups and, more recently, social listening approaches. In contrast to ‘active’ market research techniques, the latter describes a set of methods that rely on existing data such as consumer dialogues and product reviews in social media to extract jobs-to-be-done, need gaps and pain points (Roberts/Piller 2016). While customer data are comparatively easy to access via market research and have proved to be a valuable source of incremental innovation, companies may run the risk of missing out on more disruptive innovation potentials. On this account, new approaches were introduced such as the lead user method, the concept of trend receivers or emergent nature consumers (Hofmann 2015) directing attention to special breeds of users with rare characteristics enabling them to better describe future, more radical needs and indicate innovation opportunities.

Dr. Volker Bilgram

Managing Director at HYVE the innovation company and affiliated researcher at RWTH Aachen University.
volker.bilgram@hyve.net

Oliver Gluth

Head of Social Media Research at HYVE with a focus on text mining and machine learning.
oliver.gluth@hyve.net

Prof. Dr. Frank Piller

heads the technology and innovation management group at RWTH Aachen University and is a co-director of the RWTH TIME Research Area.
piller@tim.rwth-aachen.de

Despite their strengths in creating empathy and understanding for future consumers and uncovering disruptive innovation potentials, targeting these special users confronts researchers with significant challenges. First and foremost, scholars point out that these rare subjects are very difficult to identify and recruit and therefore the search costs can be significant. Second, the innovation opportunities generated with usually small-N samples of special users are rather qualitative in nature and thus lack robustness and certainty. Moreover, the data collected via these approaches describes special users’ assessments and interpretations of future developments and does not necessarily represent behavioral intentions. These downsides of current methods for opportunity identification oftentimes prevent companies from allocating significant investments.

Exploring New Forms of Data-Driven Opportunity Identification

To fix this innovation conundrum, the case presented in this paper investigates a new approach for identifying and assessing innovation opportunities. In particular, our study directs attention to both new data sources and new forms of data

analysis which promise more robust prediction due to large sample sizes and a 'hard currency' to quantify innovation potential. Drawing on a unique data set of 211,899 data points from 5,727 projects in English language from major crowdfunding platforms, the study is the first to provide insights into how crowdfunding data can be used to identify innovation opportunities currently explored by entrepreneurs and backed by innovative consumers. The paper shows how this data can be analyzed by applying latent Dirichlet allocation (Blei et al. 2003) in order to aggregate and cluster innovation opportunities based on topic similarity. We further provide insights into how the funding volume allocated by consumers may indicate the attractiveness of innovation opportunities.

Adequate Means of Analysis – the Rise of Machine Learning and Artificial Intelligence

As traditional methods of analysis fall short of managing the ever-increasing volumes of unstructured data, innovation researchers pin their hope on artificial intelligence (AI). In particular, the emergence of powerful AI engines and open repositories of algorithms gave rise to automated analyses via machine learning. Machine learning makes use of efficient algorithms capable of recognizing, structuring and organizing innovation-related information such as consumer preferences or product features from large-scale unstructured datasets (Tirunillai/Tellis 2014). Christensen et al. (2016), for instance, used a supervised machine learning technique to train a model to detect ideas in online communities. Within the domain of machine learning, topic modeling is a frequently used text mining approach enabling researchers to uncover hidden useful structures in otherwise unmanageable large text collections (Blei et al. 2003). The algorithms applied for topic modeling can help to identify and define previously unknown topics, i.e. a collection of words with a relatively high usage probability, in single text documents. Based on these topics, statistical procedures can assess the similarity between the text units in the collection and reveal clusters without any manual pre-categorization.

Latent Dirichlet allocation (LDA) is a prevalent topic modeling algorithm first introduced by Blei et al. (2003). In contrast to supervised learning techniques (e.g. neural networks or support vector machines), LDA is an unsupervised learning technique aiming at detecting patterns in unstructured data by identifying and explaining similar data based on their distribution on latent topics. LDA relies on a bag-of-words assumption, i.e. the model only considers word co-

occurrences as opposed to the relative position to one another (Antons et al. 2016). Applying LDA can be used to identify topics within text corpora for various purposes such as brand analyses, sales prediction or brand positioning (Tirunillai/Tellis 2014; Lee/BradLow 2011).

Crowdfunding Data – the New Kid on the Block

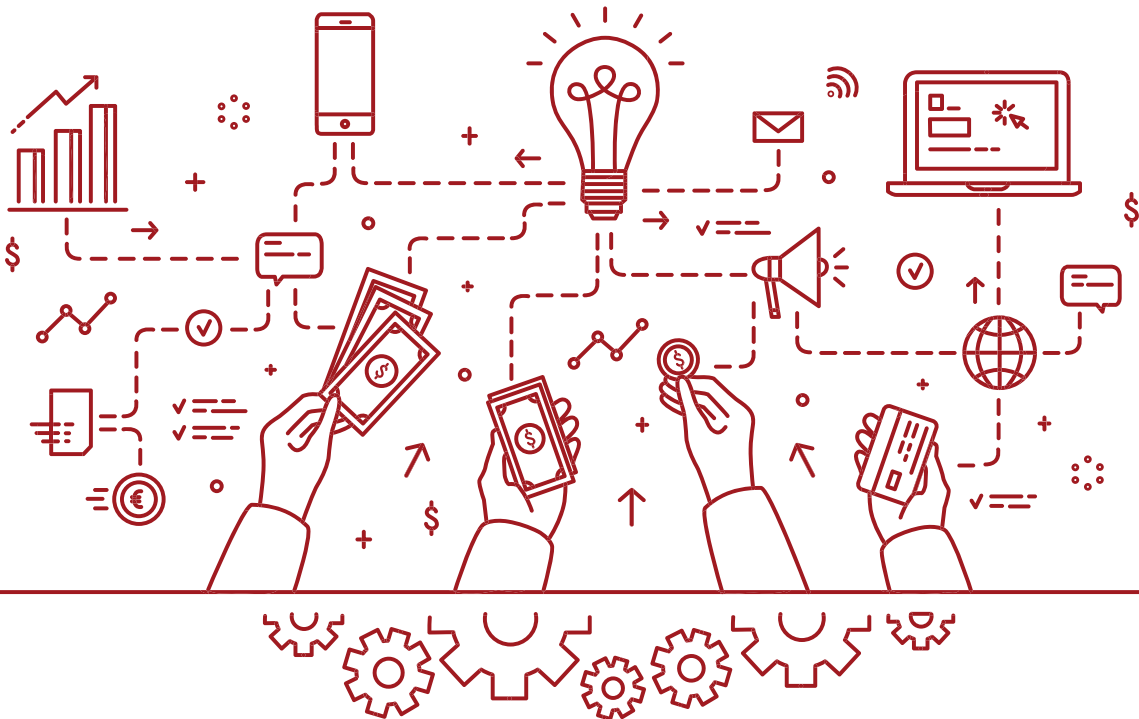
A special case of social media sites providing powerful consumer data to deduce innovation opportunities from are crowdfunding sites. Based on Schwienbacher and Larralde (2010), crowdfunding is defined as an open call, essentially through the Internet, to raise financial resources for specific purposes from a large number of capital providers (i.e. the crowd) in exchange for some form of reward.

With regard to identifying innovation opportunities, the concept of crowdfunding is particularly notable due to a distinctive feature. In contrast to traditional modes of funding, which create the product first before getting a market reaction, crowdfunding offers an early feedback mechanism at the prototype stage. Thereby, the crowd investors essentially buy a product before it is produced based only on an early prototype or a mere concept. As opposed to traditional market research measuring mere behavioral intentions to buy a product, consumers funding projects on crowdfunding sites actually invest real money. Thus, crowdfunding data provides a 'hard currency' for the attractiveness of the product and shifts feedback mechanisms to early phases of new product development. In other words, crowdfunding platforms are

Management Summary

In the digital era, a plethora of unstructured (consumer) data is available to companies to unlock innovation opportunities. Among these data sources are crowdfunding platforms, which serve as the litmus test for the desirability and viability of thousands of innovative ideas today. Based on the unique data on consumers' actual purchasing behavior, crowdfunding data bears the potential to indicate future innovation opportunities. The machine learning approach introduced in this study shows how companies can use artificial intelligence to identify innovation opportunities from crowdfunding data in an automated fashion.

CROWDFUNDING



large organized marketplaces systematically matching innovative entrepreneurs with a need for capital and individuals with an unanswered need. In doing so, we argue that crowdfunding platforms serve as markets before the mass market and may forecast market opportunities without the necessity to undergo capital intensive production processes.

Despite the recent surge of attention to the phenomenon, the use of crowdfunding data to identify innovation opportunities and weak signals for future markets remains largely undiscussed. This is particularly surprising considering popular and highly successful crowdfunding initiatives that raised millions of dollars and quite often paved the way for entirely new markets. The Pebble smartwatch as well as smart home gadgets by SmartThings are frequently cited examples of crowdfunding-born products that not only introduced a product innovation but heralded entire new product categories significantly prior to market launches of bigger players. Although one might argue that the original opportunity has already been seized by another party once the idea is publicly crowdfunded, analyzing campaigns relevant to their business bears significant options for companies. First, companies may consider buying a company, especially, when technological expertise and intellectual property is involved which is hard to copy. The acquisition of Oculus Rift by Facebook serves as an example of this strategy. Second, companies may realize the potential of a product idea that resonates well with consumers on crowdfunding sites

The analysis of crowdfunding data is powerful as it reflects actual purchase behavior and indicates future revenue.

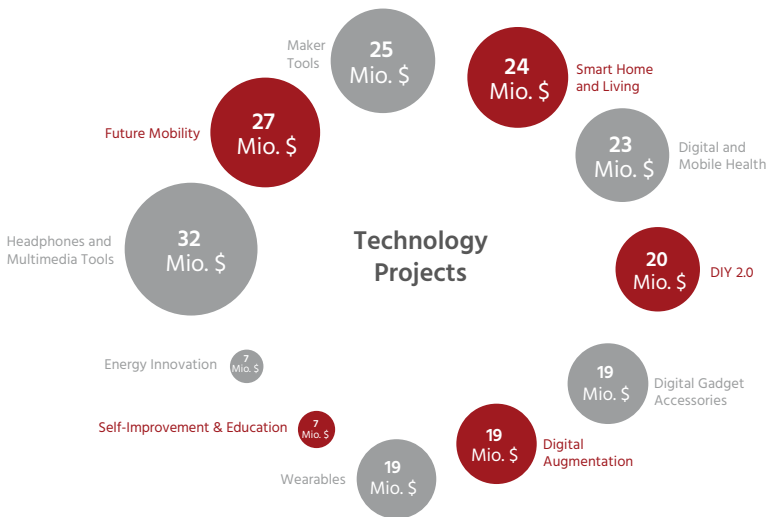
and copy the idea. In light of the oftentimes lengthy and precarious production processes as well as issues regarding delivery following a successfully funded campaign, companies may have the edge over small startups on crowdfunding sites due to their established processes and capabilities (Mollick 2014). Despite the time lag, companies may still produce their own product innovation, scale faster and outdo the competition in terms of speed and quality.

Machine Learning – Data Retrieval and Topic Extraction

In a first step, we retrieved data from the websites of leading crowdfunding sites such as Kickstarter using a crawling mechanism that extracted the following information from crowdfunding projects:

- description of project
- funding goal, i.e. the goal set by the entrepreneur for the specific project

Fig. 1: Meta Cluster of Technology Projects



Source: 2015 HYVE AG.

- achieved funding, i.e. the total amount of funding provided by backers at the end of the project period
- end date of the project

In order to limit the amount of data we focused on crowdfunding projects from the technology category, which is one

of the predefined project categories on the crowdfunding sites that entrepreneurs can choose to describe the general nature of their endeavor. Furthermore, we also confine the period, in which the crowdfunding projects took place, to the years 2013–2015. This procedure led to a total number of 5,727 technology-related crowdfunding projects and 211,899 data points.

LDA analysis was applied using the topicmodels package implementation for the statistical software R. Within the data set of 5,727 project descriptions, the machine learning procedure extracted a total of 68 distinct topics consisting of 37 to 206 single projects. It provided a distribution of probability for all projects within the data set to fall into one of the computed topics. Furthermore, for each topic the most characteristic terms were determined, i.e. the terms that are most constitutive for the respective latent topic. Based on these terms as well as on the descriptions of the most characteristic projects, the 68 automatically extracted topics were then manually described by experts and allocated to 11 larger coherent meta clusters.

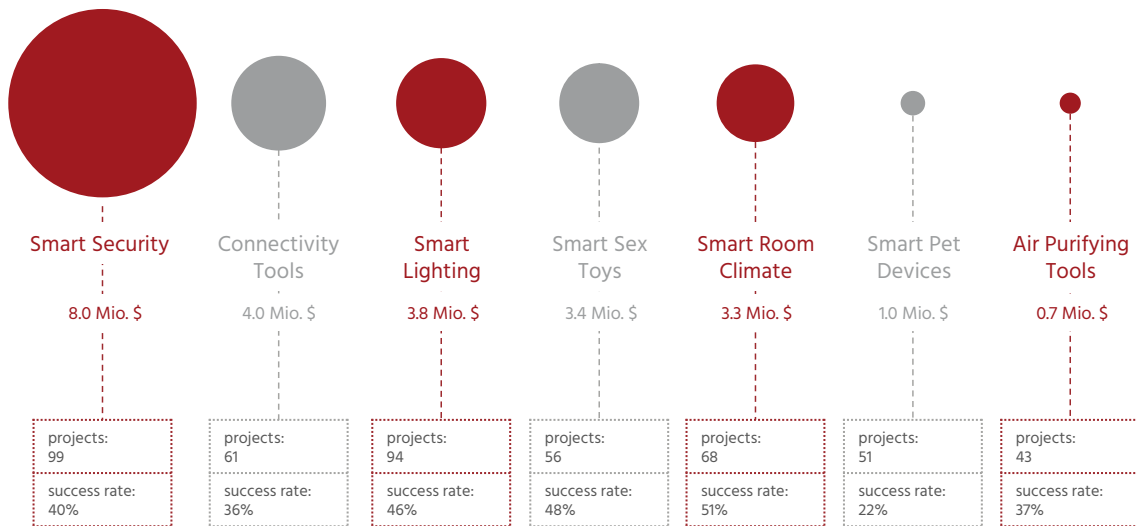
Opportunity Fields Revealed by Machine Learning

The approach provided several insights into how crowdfunding data can be utilized to identify hot topics among tech-savvy users and promising innovation opportunities. Figure 1 provides an aggregated view of 11 meta clusters which resulted from the manual labeling of the original clusters

Lessons Learned

1. Identify and structure the data sources (internal and external) that hold valuable information for your business regarding the desirability (what do consumers need and love?), the feasibility (what are technically doable solutions?) and viability (are there business models that work for consumers and your company?).
2. Clarify whether data is accessible or can be purchased - either retrieve the data by means of crawling algorithms or purchase access to the data.
3. Synchronize the need for information 'decision makers' in your organization have with the capabilities of your data scientists by shaping integral teams consisting of both groups.
4. Initiate pilot projects to explore the opportunities of new data sources and means of analysis (e.g. machine learning) and learn about their respective strengths.
5. Embrace a data-driven process for opportunity identification for new product development to find fresh and relevant innovation potentials off the beaten track.
6. Implement a data strategy empowering managers to incorporate the data layer in any business decision.

Fig. 2: Clusters Within the Smart Home and Living Meta Cluster



Source: 2015 HYVE AG.

computed via LDA analysis. The diameter of the circles represents the total amount of funding within each cluster computed by the sum of collected money of successful crowdfunding projects.

This aggregated level provides an overview of the general product categories and application fields of crowdfunding projects.

Crowdfunding data provide robust insights into key attributes of successful innovation: the feasibility, desirability and viability of new products.

ding projects in the technology segment between 2013 and 2015. Unsurprisingly many popular innovative segments and trend topics appear: Smart Home and Living, Wearables or Digital and Mobile Health are all sound fields within the public discourse of innovation management. For instance, the identified clusters show strong similarities with the technologies highlighted in the annually published Gartner

Hype Cycle for Emerging Technologies supporting the plausibility of the approach.

In contrast to the trajectories elaborated in many innovation and trend reports, which rely on expert interviews and statistical projections based on indicator data, the clusters and funding volumes in figure 1 reflect actual purchase behavior and the respective revenue. This provides a quantitative real-world reference of entrepreneurial efforts and consumer demand that can be used to assess the maturity and attractiveness of specific innovation fields. As shown in figure 1, the three clusters with the highest funding are Headphones and Multimedia Tools, Future Mobility and Maker Tools, indicating which innovation opportunities attract the most attention from both consumers, i.e. the backers of projects, and entrepreneurs, i.e. the funding seekers. For example: While Smart Home and Living and Digital and Mobile Health projects receive comparably high attention from project owners and funders, both clearly lag behind the amount of consumer investments in Headphones and Multimedia Tools.

While this highly aggregated view on the results provides a good overview of the fields of attention within the crowdfunding sphere, a detailed view of the sub-clusters allows drawing more specific conclusions for development or investment decisions within companies. Therefore, we

selected one of the aggregated cluster to exhibit the entailed information of our analysis on the next level of detail: Figure 2 exemplifies the Smart Home and Living meta cluster with its constituting subcategories and their contribution to the overall funding.

Projects within this cluster reflect the increasingly prevalent connectivity in today's households and the progressing efforts of automation in all areas of everyday live covering a range from purpose-specific smart modules to scalable smart living suites. While the relatively large funding volume of smart living features like Smart Lighting and Smart Room Climate are within expectations, the significantly larger amount of collected money from projects addressing Smart Security is a noteworthy signal for a rewarding smart home segment. In contrast to this, the category of Smart Pet Devices might be a plausible idea, but the received funding as well as the average success rate indicate a rather moderate market potential.

Conclusion

In this paper, we have described a novel method for identifying innovation opportunities drawing on data of crowd-

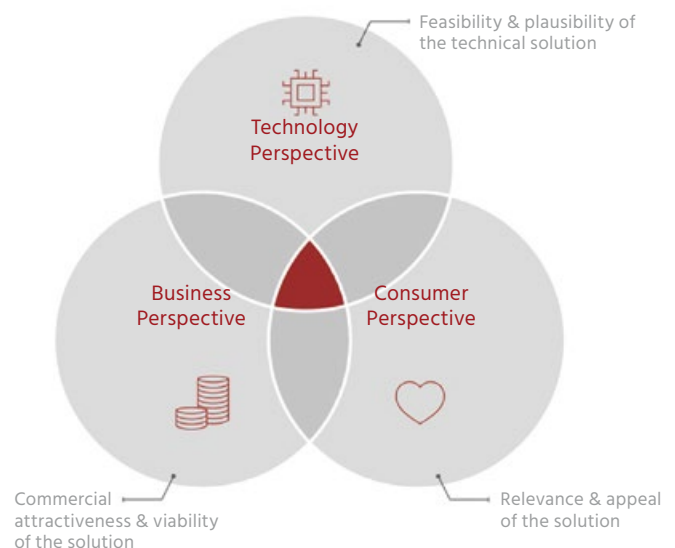
funding projects initiated by entrepreneurs and backed by innovative consumers. The approach is novel in two respects: first, it uses large datasets from a unique data source, namely crowdfunding sites. Second, it sheds light on the domain of artificial intelligence showing how machine learning, in particular LDA, can be applied to analyze the data in an automated way. Thus, we join the stream of research on automated text analysis and topic modeling via machine learning and showcase a new and unprecedented way of utilizing the unique consumer data on crowdfunding sites to identify triggers for innovation.

We argue that, due to the nature of the crowdfunding phenomenon, this data source is particularly relevant for identifying innovation opportunities. Crowdfunding sites serve as a marketplace indicating the attractiveness of new product concepts measured as the aggregated 'de facto' payments made by consumers prior to the realization of the product concept. In contrast to analyzing consumer preference data, e.g. product reviews or conversations about product offerings, the price paid by consumers to practically pre-order a future product serves as a 'hard currency'. Therefore, the data allows for a paradoxical insight – it reflects actual purchasing behavior before the product is even realized and adopted by the mass market. Moreover, crowd-

Main Propositions

1. Consumer data retrievable from social media such as crowdfunding sites are a valuable source of innovation and trends.
2. The actual purchasing behavior captured in crowdfunding data serves as a 'hard currency' for consumer acceptance and as an indicator for innovation opportunities.
3. Crowdfunding data provide robust insights into key attributes of successful innovation: the feasibility, desirability and viability of new products.
4. AI-based approaches are capable of automatically structuring and exploring patterns in huge, unstructured data pools.
5. Topic clusters automatically identified via machine learning help to monitor and prioritize key innovation opportunities.

Fig. 3: The Three Innovation Perspectives Addressed by Crowdfunding Data



Based on: Brown 2008. Source: 2015 HYVE AG.

funding data allow established companies to keep track of the dynamic developments of startups in their field of business and learn from the market reaction.


Our findings show that machine learning approaches are indeed capable of identifying sound clusters of innovative product concepts by applying LDA to crowdfunding

Our machine learning approach allows companies to identify innovation opportunities from crowdfunding data using artificial intelligence.

data. Within the dataset of 5,727 crowdfunding projects in the technology sector, we eventually arrive at 68 clusters. A sanity check of the clusters confirmed the usefulness and accurateness of the clusters derived automatically based on textual descriptions. The clusters proved to be topically coherent and sufficiently homogeneous. Besides the clusters derived by applying LDA, we also allocate the funding volume the crowdfunding projects in each cluster accumulated. In doing so, the topical clusters are enriched by an indicator for relevance, feasibility of the solution and commercial viability of the opportunity field. While it is

striking that the clusters read like a ‘who is who’ of current trend and technology reports, the convincing argument for the clusters identified via machine learning is that they were automatically derived from behavioral data, i.e. purchase behavior.

We argue that crowdfunding data, in particular the topic clusters with the corresponding funding volume, address three important perspectives of innovation opportunities more reliably than most conventional market research techniques (figure 3): (1) the consumer perspective, i.e. the relevance and appeal of the solution, (2) the technology perspective, i.e. the proof of concept and elaboration of a technically feasible and plausible concept, and (3) the business perspective, i.e. the value that can be captured by a commercial model.

Thus, the data-driven approach introduced in this paper may have the edge over conventional trend analysis and market research methods in terms of objectivity and validity as the opportunities are automatically extracted and assessed based on actual purchase behavior. While we intended to highlight that crowdfunding data can be a source of innovation, we also want to encourage scholars and practitioners to investigate further valuable data sources for opportunity identification. 

Literature

Antons, D./Kleer, R./Salge, T.O. (2016): Mapping the topic landscape of JPIM, 1984–2013: In search of hidden structures and development trajectories, in: *Journal of Product Innovation Management*, 33, 6, pp. 726–749.

Bilgram, V./Füller, J./Bartl, M./Biel, S./Miertsch, H. (2013): Eine Allianz gegen Flecken, in: *Harvard Business Manager*, 3, pp. 62–68.

Blei D.M./Ng, A.Y./Jordan, M.I. (2003): Latent Dirichlet allocation, in: *Journal of Machine Learning Research*, 3, pp. 993–1022.

Boudreau, K./Lakhani, K.R. (2013): Using the crowd as an innovation partner, in: *Harvard Business Review*, 91, pp. 61–69.

Brown, T. (2008): Design thinking, in: *Harvard Business Review*, 6, pp. 84–92.

Brunswick, S./Bilgram, V./Füller, J. (2016): Taming wicked civic challenges with an innovative crowd, in: *Business Horizons*, <http://dx.doi.org/10.1016/j.bushor.2016.11.001>.

Christensen, K./Norskov, S./Frederiksen, L./Scholderer, J. (2016): In search of new product ideas: Identifying ideas in online communities by machine learning and text mining, in: *Creativity and Innovation Management*, <http://dx.doi.org/10.1111/caim.12202>.

Hofmann, R. (2015): Visionary competence for long-term development of brands, products, and services: The trend receiver concept and its first applications at Audi, in: *Technological Forecasting and Social Change*, 101, pp. 83–98.

Jeppesen, L.B./Lakhani, K.R. (2010): Marginality and problem-solving effectiveness in broadcast search, in: *Organization Science*, 21, 5, pp. 1016–1033.

Lee, T. Y./Bradlow, E.T. (2011): Automated marketing research using online customer reviews, in: *Journal of Marketing Research*, 48, 5, pp. 881–894.

Mollick, E. R. (2014): The dynamics of crowdfunding: An exploratory study, in: *Journal of Business Venturing*, 29, 1, pp. 1–16.

Roberts, D.L./Piller, F.T. (2016): Finding the right role for social media in innovation, in: *MIT Sloan Management Review*, 57, 3, pp. 41–47.

Schwiebacher, A./Larralde, B. (2010): Crowdfunding of small entrepreneurial ventures, in: *Handbook of entrepreneurial finance*, <https://ssrn.com/abstract=1699183>.

Tirunillai, S./Tellis, G.J. (2014): Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation, in: *Journal of Marketing Research*, 51, 4, pp. 463–479.