Company and user preferences in Open Innovation Communities through content analysis

M. R. Martinez-Torres, S. L. Toral, F. Rodriguez-Piñero Royo
University of Seville
{rmtorres,storal,fjrpr}@us.es

Abstract: *The proliferation of Web 2.0 has encouraged new forms of innovation based on crowdsourcing, leading to open innovation schemes where customers become part of the innovation processes. However, when companies migrate to an open innovation scheme, they have to decide about which ideas deserve to be adopted. The aim of this paper is illustrating the difference between company and user preferences by analyzing the content of posted ideas.*

Keywords: Open innovation, Web 2.0, Semantic indexing, Non-reactive data collection

1. Introduction

Open innovation represents an effective strategy to provide organizations with access to a wider range of ideas in the worldwide market, reducing the costs associated with R&D (Chesbrough, 2003; Huizingh, 2011). However, open innovation also requires the community involvement to succeed (Martinez-Torres, 2013). Users and consumers participate if they feel they can improve the product of services they have experience with and if they also feel their ideas have chances to be adopted. Thus, companies face the problem of deciding which ideas have enough merit or potential profitability to be adopted. If decision making is only guided by applicability and potential profitability criteria, many customers can feel frustrated as their ideas never will become a reality. But approaching too much to user preferences can mean adopting rare or too expensive ideas.

This paper aims to further research the differences between users and company preferences by means of shared content analysis. However, this analysis cannot manually performed, since open innovation platforms usually receive a huge number of ideas (Martinez-Torres et al., 2013). This paper proposes using natural language processing techniques to compare the main topics of adopted and non adopted ideas. The advantage of these techniques is that hundreds or thousands of ideas can be computationally processed. The main limitation is treating the complexity of natural language using a computational algorithm, which obviously is a simplification.

The rest of the paper is structured as follows. Section II introduces open innovation communities. Section III describes the case study and the methodology based on semantic techniques. Obtained results are shown in Section IV as well as the discussion. Finally, Section VI concludes the paper.

2. Open Innovation Communities

One of the most popular alternatives for open innovation implementation is open innovation communities. They promote the generation of new ideas, the interactions among users as well as the interactions among the development team and customers (Di Gangi and Wasko, 2009). Interactions among users enable them to build on one another's knowledge and experiences, which plays a critical role in developing ideas. Besides, emerged discussions from posted ideas also contributes to concept testing through the comments posted by other users or through a scoring system. However, the practical implementation of open innovation communities demonstrates that they tend to generate a huge volume of information that can be difficult to manage.

This paper provides a procedure to evaluate to what extent a company following an open innovation strategy is listening to the community, and in which specific areas this is happening more intensely.

3. Case study and Methodology

3.1 MyStarbucksIdea

Starbucks is a company that pursues to satisfy a traditional necessity in a different manner. The distinctive element of this company in respect to the competitors is to offer its clients a quality service at all levels. The open innovation website is actually a fundamental element in the strategy. Through the "My Starbucks Idea" website, users can not only post and share ideas with the rest of users, but they can also comment and vote other previously posted ideas.

3.2 Data collection

Starbucks' open innovation website identifies members' contributions as ideas. When posting an idea, registered users must choose one of fifteen subcategories that respond to three basic aspects of the company: product, experience and involvement ideas, Table 1. To perform a content analysis, the header of posted ideas have been collected using a crawler (Martinez-Torres et al., 2013). The number of collected ideas per category is shown in Table 1, third column. The header is a summary of the content and usually contains the keywords or the more relevant terms related to the ideas' content.

	Category	**Nº**		**Category**	**Nº**
Product ideas	Coffee & Espresso Drinks	9500	**Experience ideas**	Ordering or Payment & Pick-Up	6338
	Frappuccino & Beverages	2687		Atmosphere & Locations	9500
	Tea & other drinks	7405		Other Experience Ideas	9500
	Food	9500	**Involvement ideas**	Buiding Community	4453
	Merchandise & Music	7113		Social Responsability	6984
	Starbucks Card	9500		Other Involvement Ideas	4699
	New Technology	2633		Outside USA	1208
	Other Product Ideas	8508			

Table 1. Categories and subcategories of posted ideas at MyStarbucksidea.

3.2 Content analysis

Natural language processing (NLP) is a set of techniques from a subspecialty of computer science and linguistics that uses computer algorithms to analyze human (natural) language. The simplest approach to deal with text analysis consists of obtaining the term-document incidence matrix, where each cell contains the number of times each word appears in each document. However, the high dimensionality of the resulting feature space is a problem when working with big collection of documents. Therefore, it is desirable to first project documents into a lower-dimensional subspace in which the semantic structure of the document space becomes clear. For instance, Latent Semantic Indexing (LSI) decomposes a term document matrix using a technique called singular value decomposition to construct new features as combinations of the original features, significantly reducing the high-dimensionality problem of the feature space (Deerwester et al., 1990).

4. Results

LSI was applied to the fifteen collection of documents, each one containing the headers of the corresponding ideas shown in Table 1. The selection of terms is done considering the number of occurrences, but avoiding prepositions and non-related terms. The singular value decomposition is applied preserving the 80% of data variance and, in this new reduced space, a cluster analysis is applied to obtain the relationships among categories.

The dendrogram of Figure 1 shows that the semantic organization of subcategories is different to the one proposed by the company. First, there is a clear relationships among the drinks and food offered by Starbuscks, as it can be appreciated at the lower part of the dendrogram. The upper part of the dendrogram shows the other half of product ideas which refers to Merchandise & Music and Other Product Ideas. However, Starbucks Card and New Technology appears to be more semantically close to Experience Ideas. These two subcategories together with Ordering or Payment & Pick-Up constitute the core of users' experiences at Starbucks. That means customers consider Starbucks Card and technology not as products but as facilitators of their experience. The other half of Experience Ideas is Atmosphere & Locations, which appears in the dendrogram close to Involvement Ideas. The last group showed in the middle part of the dedrogram is the group of involvement ideas, that also embraces Other Experience Ideas, despite this is a subcategory belonging to Experience Ideas.

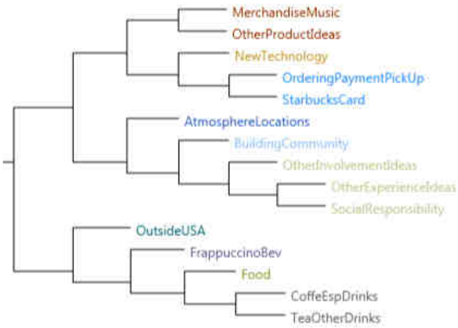

Figure 1. Categories clustered by word similarity.

Figure 2. Tag cloud for Starbucks drinks and food.

Figure 3. Tag cloud for Starbucks Merchandise & Music.

The semantic clustering shows that users interpretation of subcategories is not the same than the company original idea. Although product ideas have many subcategories in common with the original classification by Starbucks, this is not the case for Experience and Involvement Ideas. It is interesting to note that experience is clearly distinguished from the products offered by Starbusks, and it is more related to all the activities surrounding having coffee or food, like the way of ordering, payment methods, etc. New technology for instance is considered from the perspective of facilitating these activities rather than as an entertainment while sitting. On the contrary, Atmosphere and Locations is closer to the involvement of customers. Finally, Involvement Ideas is more or less similar to the original classification from Starbucks. The main difference is that Other Experience Ideas are interpreted by customers as experiences for being involved.

Figure 4. Tag cloud for Starbucks Involvement ideas.

Figure 5. Tag cloud for Starbucks Experience ideas.

Considering the dendrogram of Figure 1, the main topics within each group has been obtained in the form of a tag cloud. Figure 2 illustrate the tag cloud for the group Coffee & Espresso Drinks, Frappuccino & Beverages, Tea & other drinks and Food. It can be seen that the main topics are clearly all the different drinks and food offered by Starbucks. The second group of Merchandise & Music is shown in Figure 3. This tag cloud contains the typical cups and mugs offered by Starbucks as well as other gifts.

In general, customers clearly distinguished between products offered to be consumed from others that are part of merchandising. Some words like coffee appears in almost all groups of ideas. Therefore, it is not a word that can discriminate among groups. Involvement ideas are shown in Figure 4. They are focused on the community of users and their role as customers. Topics include recycling as the main theme related to social responsibility as well as the way in which the innovation community is working. Finally, Figure 5 corresponds to the tag cloud for Experience Ideas, where users demand innovations about ordering, cards, app and rewards. The category of new technologies which is part of this group usually refers to app and mobiles phones.

5. Conclusions

This paper set a methodology to compare company and user preferences in open innovation communities using semantic indexing techniques. Obtained results provide new insights about the innovation policies of companies, and they can be used to analyze and correct possible biases in their innovation policies.

Acknowledgements

This work has been supported by the Consejería de Economía, Innovación, Ciencia y Empleo (Research Project with reference P12-SEJ-328).

Bibliography

Chesbrough, H. (2003), Open innovation-the new imperative for creating and profiting from technology, Harvard Business School Press, Boston.

Huizingh, E. K. R. E. (2011), Open innovation: State of the art and future perspectives, Technovation, 31 (1): 2-9.

Martinez-Torres, M. R. (2013), Application of evolutionary computation techniques for the identification of innovators in open innovation communities, Expert Systems with Applications, 40 (7): 2503-2510.

Di Gangi, P. M., and Wasko, M. (2009), Steal my idea! Organizational adoption of user innovations from a user innovation community: A case study of Dell IdeaStorm, Decision Support Systems 48 (1), 303–312.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990), Indexing by latent semantic analysis, Journal of the Am. Society of Information Science, 41 (6): 391–407.