

The propensity to exploration and exploitation strategy: from theory to content analysis application

Paola Pisano, Marco Pironti
 E-business Lab
 University of Turin
 Turin, Italy
 e-mail: pisano@di.unito.it, pironti@di.unito.it

Michele Simoni
 Department of Business Studies
 Parthenope University of Naples
 Naples, Italy
 e-mail: michele.simoni@uniparthenope.it

Mauro Girauda
 Department of Computer Science
 University of Turin
 Turin, Italy
 e-mail: girauda@di.unito.it

Gianluca Vagnani
 Department of Management and Technologies
 Sapienza, University of Rome
 Rome, Italy
 e-mail: gianluca.vagnani@uniroma1.it

Abstract – The objective of this work is to investigate the content analysis methodology in the strategic propensity such as explorative/exploitative behavior. After theoretical analysis on the meaning and definition of exploitation and exploration, the authors have applied the content analysis methodology to a practice case. The conclusion is that the qualitative content analysis is a good methodology to study the strategy propensity having the same results of the other methodologies used in the strategy researches such as practice case and grounded theory.

Keywords-component: content analysis; SVM; exploration; exploitation.

I. INTRODUCTION

The panorama of the different methodologies applied to the strategy studies is really vast. Three of these are being particularly used Action research, Case studies and Grounded theory. The table 1 summarize the use of these methodologies on the strategic problematic in the literature researches.

In the action theory the researcher has to work in a close contact with the organization to define the strategies. In the Case studies, a team or a singular researcher carry out the profound work in the analysis of the organization strategic problematic, using written documents, interviews and analysis of the data. At the end in the Grounded theory the researcher formulate the theory useful for the subject involved in the analysis, through the reality [7]. In this framework the content analysis is a new methodology that have provided good results in the strategy studies. The content analysis is a “systematic technique, replicable for the compression of huge number of worlds in a define number of content categories on the basis of explicit codify rule” [29].

In this work the author will applied the content analysis to define the exploration/exploitation propensity in the organization.

TABLE I. THE USE OF DIFFERENT METHODOLOGIES IN STRATEGIC PROBLEMATIC.

Journal	Grounded Theory	Action research	Case studies
Administrative Science Quarterly	14	0	19
Journal of Business Ethics	56	19	118
Journal of Organizational Behavior	11	8	20
Management Science	5	2	42
MIS Quarterly	14	11	40
Public Administration Review	6	10	85
Risk Management	4	3	13
Strategic Management Journal	16	1	63
The Academy of Management Journal	22	13	20
The Journal of the Operational Research Society	13	50	201

After a content analysis description and the definition of the exploration and exploitation strategy, a new qualitative content analysis methodology is applied to a practice case demonstrating the goodness of the methodology.

II. THE CONTENT ANALYSIS

The content analysis born in the 1931 and became famous thank to “The Constant Comparative Method of Qualitative Analysis” [11] in which the content analysis was defined like a “Grounded Theory” extension. The content analysis bases its structure on the text to analyze, on framework in which the text is inserted and on definition of the concept to investigate [17]. The different application of the content analysis on strategic problematic have been investigated on different journal from 1997 to 2010 [1][2][5][8][14][23][24][28][30][31][33].

The content analysis can be approached in quantitative or qualitative way.

The quantitative approach analyses the concept by the frequencies of a set of word (more high is the frequency of the words in the words list more stress is the concept inside the text). In the quantitative content analysis the “word crunch” and the “Categories Tagging” are the method most used.

The principal different between the two approaches is that in the word crunch the categories are defined using just one set of words (or dictionary) instead of more dictionaries using in the Category tagging.

While the quantitative approach transforms the categories in statistic data, the qualitative ones allow to investigate on the text regardless of the word used to express the categories. The qualitative method used a combination of word or a sentence to describe the category analyzed such as happened in the natural language.

In the qualitative approach the methods more use are the Naive Bayes, Neutral network and Support Vector Machine (SVM).

III. THE DEFINITION OF THE CONCEPTS: EXPLORATION AND EXPLOITATION STRATEGY

Many studies have been done on the analysis of exploration and exploitation and their interconnection. From March [19] to Moran, Vagnani and Simoni [22], Benner and Tushman [3][4], Ghemawat and Ricart i Costa [10], Gupta, Smith, Shalley [12] and McGrath [20].

In this paper the category of the exploitation is defined as the introduction of new alternatives that born and develop thank to continuously adjustment inside the organization [26]. The alternative strategies that organization takes in consider are limited to knowledge and skill that already reside inside the company. The exploitation activities include such things as “refinement, choice, production, efficiency, selection, implementation, execution” [19]. Vice versa the exploration category considers new alternatives also out of the organization boundary, taking in consider competitors and industry where develops its businesses. Exploration activities include things like search, variation, risk taking, experimentation, play, flexibility, discovery, innovation [19]. It’s obvious that both the behaviors have a point of strength and weakness but the real difficulty is to guarantee a good balance between exploration and exploitation strategies [18].

An organization that based all strategies on the exploration excluded exploitation will be suffered of high experimentation and research costs. On the contrary an organization that based all strategies on exploitation risk to fall down in the optimization trap improving only the skills already acquired don’t taking new skills or the innovations arriving from the industry. If we compare the return of the two strategies the exploration has returns less certain and more time remote while exploitation avoid the activities with less skill that lead to a return of investment under the expectation [13].

IV. PRATICAL CASE: APPLICATION OF SVM TO THE CASE OF BAUSCH & LOMB BY SVMLIGHT

The propose of this paragraph is the application of the qualitative content analysis to Bausch & Lomb. The content analysis was done from 1989 to 2004 on the section “managerial discussion and analysis” of the balance sheet (K10) and the “letter to shareholder”.

While, the text choice is in line with the literature on the methodology (e.g.: Uotila et al. [30]), the methodology selection depend on next position: difficulty in the strategic concepts identification: the exploration and exploitation strategies couldn’t be described with a simple list of words; difficulty to separate in a clear way the two category concepts; the needs to decontextualize the strategic concepts from the industry.

A. *The company*

Bausch & Lomb is an eye health company dedicated to Bringing Visionary Ideas to Eye Health.

Bausch & Lomb traces its roots to 1853, when John Jacob Bausch, a German immigrant, set up a tiny optical goods shop in Rochester, New York. When he needed more money to keep the business going, Bausch borrowed \$60 from his good friend, Henry Lomb. Bausch promised that if the business grew, Lomb would be made a full partner. The business did grow and the partnership was formed.

Baush & Lomb bases its activities on three macro area.

Vision Care: the contact lens offerings span the entire spectrum of wearing modalities and include such well-known brand names as PureVision®, SofLens®, Boston® and Optima® .

Pharmaceuticals: the products treat a wide range of eye conditions including glaucoma, eye allergies, conjunctivitis, dry eye and retinal diseases. B&L offers proprietary and generic medicines available by prescription, over-the-counter eye drops and other medications.

Cataract and Vitreoretinal Surgery: B&L offers a full suite of products including intraocular lenses and delivery systems featuring the Crystalens®, SofPort® and Akreos™ brands of IOLs, the Stellaris® and Millennium™ lines of phacoemulsification equipment, and other surgical instruments and devices, including the Storz® line of instruments.

B. *The content analysis application*

The SVM application was done follow the next steps: software selection; definition of the coding scheme for the concepts to analyze and validity text; text coding; evaluation of the encoding consistency with the concepts definition; analysis of the results.

C. *Software selection*

The software used in the text coding was SVMlight for its efficiency in computational time and in the coding results. SVMlight belong to SVM (Support Vector Machine) a set of supervised learning methods for regression and classification of patterns. They belong to the family of generalized linear classifiers also known as maximum margin classifiers, because at the same time minimize the empirical

classification error and maximize the geometric margin, i.e. the distance of the point corresponding to an element of the training set from the hyper plane of separation. Given a training set the natural objective is to find a hyper plane separating the two classes that maximizes the geometric margin, extending as far as possible the distance between the hyper plane and the points corresponding to the elements that increase the confidence of the classification. The SVM can be thought as an alternative technique for learning polynomial classifiers, as opposed to the classical techniques of neural networks training. This type of instruments presents an efficient algorithm and is capable of representing complex nonlinear functions. The use of SVM in the literature is not new, F. Colas and P. Brazdil [6] wrote an article comparing the SVM and the Naive Bayes method, E. Tkacz and P. Kostka [16] proposed a success case study of SVM for the management of medical data, other authors analyzed how changing a few variables of the SVM could change the performance of coding, A. Khoo, Y. Marom and D. Albrecht [15] applied SVM to a case study analyzing the inclusion or not of stop word within the text, S. Shalev-Shwartz and N. Srebro [27] analyzed the relationship between the magnitude of the training set and the performance of SVM.

In the case of linearly separable sets the basic idea of SVM is to construct the separating hyper plane (optimal hyper plane) that maximizes the distance between the elements of different classes.

In this way the concepts to be analyzed (e.g., exploration/exploitation) are well separated and defined by a set of sentences. The optimal hyper plane defines a surface that separates the elements of the two classes. The property of maximize the margin between the elements of the two classes ensure to SVM good chances of generalization in classification problems.

The operation of allocating a classification to the sentence (i.e. explorative or exploitative) occurs according to the optimization of a linear SVM, especially it is to optimize the following equation:

$$\alpha^* = \text{maximise}_{\alpha} \sum_{i=1} \sum_{j=1} y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \quad (1)$$

with the limitations imposed by conditions of Karush-Kuhn-Tucker (KKT):

$$\sum_{i=1} y_i \alpha_i = 0 ; 0 \leq \alpha_i \leq C ; i = 1 \dots l \quad (2)$$

where α is the weight of the samples, C the relative importance of model complexity and error. This leads to a prediction function:

$$\phi(x') = \text{sign}(\sum_{i=1} \alpha_i y_i \langle x_i, x' \rangle + b) = \text{sign}(\langle w^*, x' \rangle + b) \quad (3)$$

where

$$w^* = \sum_{i=1} \alpha_i y_i x_i \quad (4)$$

The results of the classification using the SVM can be represented according to the figure below, where the hyperspace is simplified to a two-dimensional plane and the optimal hyper-plan in a straight line.

For each sentence is assigned a prediction based on the model built during training: a positive result identifies the phrase as belong to the category, while a negative result identifies it's not belonging to the same category.

In the figure 1 the black squares and circles represent the support vectors selected by SVM training within the collection, respectively, to define the margin positive H1 and negative H2, while the white squares represent vectors, or sentences, with positive classification, the white circles those with a negative rating. Their prediction is identified with the optimal hyper-plane distance: greater is this distance in terms of absolute value greater is the confidence that can be attributed to the classification. Tables 2 and 3 reports some examples of classification of sentences exploration and exploitation of the Bausch & Lomb in 1991.

D. Definition of the coding scheme for the concepts to analyse and validity text;

The sentence was selected as the basic units of the analysis. The basic unit definition is really important [32] because could influence the analysis results [9] and impact on the definition of the coding scheme.

As far as the coding categories, the exploration and exploitation categories are defined in SVMlight thanks to a set of sentence that defines the salient features.

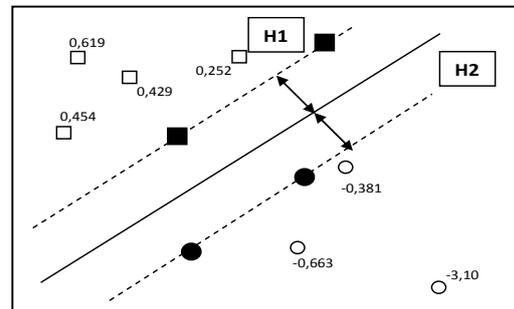


Figure 1. Visual representation of the exploit sentence with SVM.

TABLE II. EXAMPLES OF SENTENCES IN THE EXPLOIT CATEGORY

Sentence	Prediction
we are very pleased to report that the complex challenges encountered in 1991 were overcome, and that bausch & lomb had a highly successful year	-0,38159
the company manages its established core businesses for both growth and cash flow, and continues to seek opportunities for improving asset management and reducing manufacturing costs	0,61917
sales for oral care products outside the usa increased more than 70 % from the prior year based on growth in europe, canada, and the far east	-0,66395

TABLE III. EXAMPLES OF SENTENCES IN THE EXPLORE CATEGORY.

Sentence	Prediction
----------	------------

we are very pleased to report that the complex challenges encountered in 1991 were overcome, and that bausch & lomb had a highly successful year	-0,4496
our development strategies in this area encompass proprietary research, product licensing, joint ventures, and acquisitions	0,53283
we stand ready to fill this need through our new pharmaceutical plant that was completed in tampa, florida during 1991	0,32798

In the following boxes the authors selects examples of sentences that defined the explorative/exploitative behaviors.

Exploitative sentence

- The new technology requires adjustments to the design of existing models and will increase manufacturing costs in the near term, however, per unit costs are expected to decline as sales of these engines increase.

Explorative sentences

- ..., we continue to make substantial investments for the expansion of our current product lines and future products to promote strong, long-term growth.

The number of sentences must be consistent in order to cover all the different nuances of the concept to analyze. By the way, there is no “golden rule” that indicates the amount of sentences necessary to define adequately the training set. The parameter that indicates the goodness of all training is the effectiveness of the classification model built assessed on a test set belonging to the corpus. Define a concept through a set of words rather than sentences are a difficult task especially in terms of consistency. It is therefore useful to work both sessions of separate coding and coding joint sessions of the resources allocated by the project. In the practice case pre-classified examples constitute the training set for SVM algorithms, that through the procedures of optimization used, in this case SVMlight, will lead to the development of a classification model that can be applied to the entire corpus. The coding model is tested on a portion of the text, and based on the result it’s possible to proceed to optimize the algorithm parameters.

E. Text coding

When sufficient consistency has been achieved, the coding rules can be applied to the entire text. During the coding process, it’s useful to check the coding repeatedly, to prevent “drifting into an idiosyncratic sense of what the codes mean” [25]: in this way it must avoid the risk of building words and sentence structures in accordance with the imagination and their (personal) cognitive structure to tell the coding as we would.

This is the step with more use of machine time, since the entire corpus of data must be encode to the software procedure. Human intervention at this point is much reduced, since you may use software procedures which carry out the work themselves and produce the results of the classification.

F. Evaluation of the ecoding consistency with the concepts definition.

After coding the entire data set, you need to recheck the consistency of the codify. It is not safe to assume that, if a sample was coded in a consistent and reliable way, the coding of the whole corpus of text is consistent as well. New

codes may have been added since the original consistency check, the understanding of categories and coding rules may change subtly over the time, which may lead to greater inconsistency [21][32]. This step is one that involves the greatest amount of work by the research team, as we must re-check the results of coding, correct dictionaries or training set and then iteratively repeats the steps outlined so far, to obtain a correct and consistent coding.

G. Analysis of the results

This step is to make explicit the meaning of the themes and categories identified and their properties. The activities may involve exploring the properties and dimensions of categories, identifying relationships between categories, discovering new categories and models to be applied to the full range of data. The table 4 sums up the results of coding, i.e. the numbers of sentences coded as exploit and explore the proportion of total sentences in the text and finally the degree of exploration of an enterprise during the reference year that was measured by the following formula:

$$\text{Degree of Explor} = \frac{|\text{Explor}|}{(|\text{Explor}| + |\text{Exploit}|)} \quad (5)$$

Same data have been reported in the figure 2.

V. CONCLUSION

Analyzing the graph of content analysis on exploration and exploitation outlined it possible to summarize the next results.

High degree of exploitation: in 1990 0,69% of exploitation against 0,31 of exploration; from 1992 to 1994: an average of 0,63% of exploitation against 0,37% of exploration.

High degree of exploration: in 1998: 0,61% of exploration against 0,39% of exploitation; in 2003 : 0,58% of exploration against 0,42 of exploitation.

In the analysis of these periods based on B & L practice case analysis it possible to highlight that: in 1990: B & L decided to focus its activities on core competencies of the company continuing to invest in eye care; in 1992: build a new plant for production of contact lenses and lens cleaning solutions in India. In 1998: through the acquisition of Chiron Vision and Storz Ophthalmics enter in the new market of surgery for cataract operations. In 2003: includes in its portfolio of business new activities such as property and the production of generic medicines such as PreserVision Ocular Vitamins.

The content analysis has created a methodological support for the definition of strategic concepts of exploration and exploitation that is correct and in line with the strategic activities of the company in that years.

TABLE IV. RESULT OF MASSIVE DATA CLASSIFICATION OF B&L

Year	Total	Explor	Exploit	Sentence Explor rate	Sentence Exploit rate	Degree of Explor
1989	225	12	15	5%	7%	0,44
1990	267	11	25	4%	9%	0,31
1991	332	18	18	5%	5%	0,50

1992	310	12	19	4%	6%	0,39
1993	373	12	22	3%	6%	0,35
1994	486	18	32	4%	7%	0,36
1995	448	33	40	7%	9%	0,45
1996	365	37	39	10%	11%	0,49
1997	322	21	25	7%	8%	0,46
1998	361	33	21	9%	6%	0,61
1999	329	20	16	6%	5%	0,56
2000	459	39	29	8%	6%	0,57
2001	459	29	32	6%	7%	0,48
2002	629	28	31	4%	5%	0,47
2003	637	49	36	8%	6%	0,58
2004	574	46	37	8%	6%	0,55

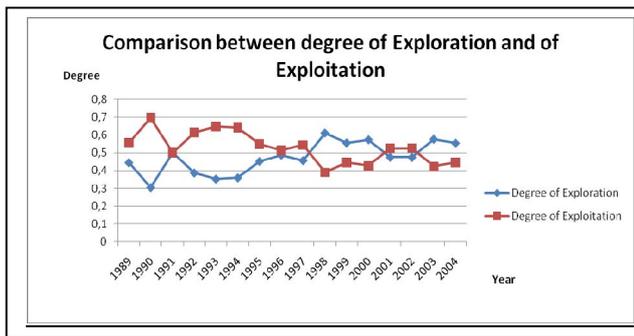


Figure 2. Chart comparing the degree of exploitation and of exploration.

REFERECES

[1] E. Abrahamson, and D. C. Hambrick, "Attentional homogeneity in industries: the effect of discretion", *Journal Of Organizational Behavior*, vol. 18, 1997, pp. 513-532.

[2] P. H. Bam, "When private beliefs shape collective reality: the effects of beliefs about coworkers on group discussion and performance", *Management Science*, vol. 49, no. 6, Jun. 2003, pp. 801-815.

[3] M. J. Benner, and M. L. Tushman, "Process management and technological innovation: a longitudinal study of the photography and paint industries", *Administrative Science Quarterly*, vol. 47, 2002, pp. 676-706.

[4] M. J. Benner, and M. L. Tushman, "Exploitation, exploration, and process management: the productivity dilemma revisited", *Academy of Management Review*, vol. 28, 2003, pp. 238-256.

[5] B. K. Boyd, S. Gove, and M. A. Hitt, "Construct measurement in strategic management research: illusion or reality?", *Strategic Management Journal*, vol. 26, No. 3, Mar., 2005, pp. 239-257.

[6] F. Colas, and P. Brazdil, "Comparison of svm and some older classification algorithms in text classification tasks", in *IFIP International Federation for Information Processing, Volume 217, Artificial Intelligence in Theory and Practice*, ed. M. Bramer, Boston: Springer, 2006, pp. 169-178.

[7] J. Collis, and R. Hussey, *Businnes research: a pratical guide for undergraduate and postgraduate students*. 2nd edition, Basingstoke: Palgrave Macmillan, 2003.

[8] R. D'Aveni, and I. MacMillan, "Crisis and the content of managerial communications: a study of the focus of attention of top managers in surviving and failing firms", *Administrative Science Quarterly*, vol. 35, 1990, pp. 634-657.

[9] B. De Wever, T. Schellens, M. Valcke, and H. Van Keer, "Content analysis schemes to analyze transcripts of online asynchronous discussion groups: a review", *Computer & Education*, vol. 46, 2006, pp. 6-28.

[10] P. Ghemawat, and J. E. Ricart i Costa, "The organizational tension between static and dynamic efficiency", *Strategic Management Journal*, vol. 14, Winter 1993, pp. 59-73.

[11] B. G. Glaser, "The Constant Comparative Method of Qualitative Analysis", *Social Problems*, vol. 12, no. 4, Spring 1965, pp. 436-445.

[12] A. K. Gupta, K. G. Smith, and C. E. Shalley, "The interplay between exploration and exploitation", *Academy of Management Journal*, vol. 49, 2006, pp. 693-706.

[13] S. R. Herriott, D. Levinthal, and J. G. March, "Learning from experience in organizations", *The American Economic Review*, vol. 75, no. 2, Papers and Proceedings of the Ninety-Seventh Annual Meeting of the American Economic Association, May 1985, pp. 298-302.

[14] R. Hodson, "Organizational trustworthiness: findings from the population of organizational ethnographies", *Organization Science*, vol. 15, no. 4 (Jul. - Aug. 2004), pp. 432-445.

[15] A. Khoo, Y. Marom, and D. Albrecht, "Experiments with sentence classification", in *proc. of the Australasian Language Technology Workshop (ALTW'06)*, Sydney, Australia, December 2006, pp. 18-25.

[16] P. Kostka, and E. Tkacz, "Support vector machine classifier with feature extraction stage as an efficient tool for atrial fibrillation detection improvement", *Computer Recognition Systems*, vol. 2, 2008, pp. 356-363.

[17] K. Krippendorff, *Content analysis: an introduction to its methodology*. 2nd edition, Thousand Oaks, CA: Sage, 2004.

[18] D. A. Levinthal, and J. G. March, "A model of adaptive organizational search", *Journal of Economic Behavior and Organization*, vol. 2, 1981, pp. 307-333.

[19] J. G. March, "Exploration and exploitation in organizational learning", *Organization Science*, vol. 2, 1991, pp. 71-87.

[20] R. G. McGrath, "Exploratory learning, adaptive capacity, and the role of managerial oversight", *Academy of Management Journal*, vol. 44, 2001, pp. 118-131.

[21] M. Miles, and A.M. Huberman, *Qualitative Data Analysis*. Thousand Oaks, CA: Sage Publications, 1994.

[22] P. Moran, M. Simoni, and G. Vagnani, "Becoming the best: by beating or ignoring the best? Toward an expanded view of the role of managerial selection in complex and turbulent environments", *Journal of Management and Governance*, in press, 2010, DOI: 10.1007/s10997-010-9129-2.

[23] R. Nag, D. C. Hambrick, and M.-J. Chen, "Is strategic management, really? Inductive derivation of a consensus definition of the field", *Strategic Management Journal*, vol. 28, no. 9, Sep. 2007, pp. 935-955.

[24] J. D. Osborne, C. I. Stubbart, and A. Ramaprasad, "Strategic groups and competitive enactment: a study of dynamic relationships between mental models and performance", *Strategic Management Journal*, vol. 22, no. 5, May 2001, pp. 435-454.

[25] J. Schilling, "On the pragmatics of qualitative assessment: designing the process for content analysis", *European Journal of Psychological Assessment*, vol. 22, no. 1, 2006, pp. 28-37.

[26] J. A. Schumpeter, *The Theory of Economic Development*. Cambridge, Mass: Harvard University Press, 1934 [Translated by Redvers Opie from the second German edition, 1926].

[27] S. Shalev-Shwartz, and N. Srebro, "SVM optimization: inverse dependence on training set size", in *proc. of the 25th international conference on Machine learning*, Helsinki, Finland, Jul. 05-09, 2008, pp. 928-935.

[28] J. C. Short, and T. B. Palmer, "Organizational performance referents: an empirical examination of their content and influences", *Organizational Behavior and Human Decision Processes*, vol. 90, 2003, pp. 209-224.

[29] S. Stemler, "An overview of content analysis", *Practical Assessment, Research & Evaluation*, vol. 7, no. 17, 2001.

[30] J. Uotila, M. Maula, T. Keil, and S.A. Zahra, "Exploration, exploitation, and financial performance: analysis of S&P 500

corporations”, *Strategic Management Journal*, vol. 30, 2009, pp.221-231.

- [31] J.P. Walsh, K. Weber, and J.D. Margolis, “Social issues and management: our lost cause found”, *Journal of Management*, vol. 29, no. 6, 2003, pp. 859-881.
- [32] R. P. Weber, *Basic Content Analysis*. 2nd ed., Newbury Park, CA: Sage, 1990.
- [33] M. S. Yadav, J. P. Prabhu, and R.K. Chandy, “Managing the future: CEO attention and innovation outcomes”, *Journal of Marketing*, vol. 71, 2007, pp. 84-101.