Text Mining project for Winter Olympic Games: Turin 2006

Veronica Baldisserri, Gianluca Bo, Flavio Bonifacio
METIS Ricerche S.r.l – Turin – Italy

1. Acknowledgements

2. Introduction

Text Mining is the process of analyzing text to extract information for specified purposes.

This paper look at a text classification task exploring attitudes of citizens towards the Winter Olympic Games that will be held in Turin in the winter of 2006.

The challenge is to understand if inhabitants of a large industrial city (about one million people) are favourably disposed towards the Games and the new image that the Winter Olympic Games will bring to Turin (*the capital city of the Alps*).

Data under examination consist of a few closed questions and an open one collected using a survey carried out with SAS tools (SAS for CATI). Text analysis is performed by means of correspondence analysis of a lexical table (Lebart, Salem, 1997), carried out with SAS tools. Results consist of a classification of interview responses in order to analyse how Turin citizens describe their city and consequently to illustrate the acceptance or refusal of Winter Olympic Games project. The outcome could be important for Public Administration to drive information campaign in order to reach the largest popular support for the Olympic event.

3. Research design

As well known, in 2006 Turin will hold the Winter Olympic Games and since 2001 Metis Ricerche, in cooperation with Turin University, has been monitoring citizens attitude towards the Games by means of specific survey CATI (a first study was presented at SEUGI20[1] in Paris in 2002).

Of course the Games represent a great opportunity for the city and studies about previous Olympic games stated that a deep rooted involvement of citizens ensures a satisfactory and enduring outcomes for the Olympic project.

Therefore Data Mining instruments are essential to help local administrations to single out those inhabitants who don't completely agree with the project, in order to structure a specific communication campaign addressed to this group.

Firstly Data Mining techniques are used here to identify this target. Here we will use respondents' answers to closed questions inquiring into their attitudes towards the Games (degree of approval, of pride and so on).

In a second step instruments such as Text Mining are used. Text analyzed is derived from general questions regarding the image of Turin. The goal here is to describe this target in order to understand which features and which channels could maximize the effect of a new information campaign, as well as to understand how the citizens perceive the city.

4. Step one – analysis of closed questions

The monitoring of Turin citizens shows an increasing awareness and interest in the project. Both in 2002 and in 2003 more than 95% of interviewed were quite or very favourable to the Games. They were proud of housing the games and agreed that the games

---

[1] V. Baldisserri, F. Bonifacio, *Turin Olympic Winter Games 2006. A marketing survey of local government initiatives, conducted with SAS tools*, SEUGI Paris, June 2002

were of great importance for the whole city. The population segments with the most favourable attitude towards the Games were males, those between 18 and 30 years old and those with a high level of education.

Local administration and the Organising Committee for the XX Olympic Winter Games (TOROC) should be satisfied with these results: an overwhelming majority of Turin citizens gave their assent to the Olympic project. Nevertheless, analyzing data deeply, there are variations in the overall level of agreement.

Closed questions used in this step are[2]:

Do you agree or disagree with the choice of Turin for the location of the 2006 Olympic Winter Games?
Strongly Agree
Somewhat Agree
Disagree
Don't know

Are you proud that Turin has won the international competition for hosting the Winter Olympic Games ?
Very proud
Fairly proud
Not proud
Don't know

Do you think you will be directly involved in any of the projects for the 2006 Turin Olympic Winter Games?
Certainly involved
Probably involved
Not involved

Do you think the 2006 Turin Olympic Winter Games will bring you any advantages or disadvantages?
Advantages
Disadvantages
Neither advantages nor disadvantages
What sort of impact do you think the 2006 Turin Olympic Winter Games will have on the region where you live?
Very positive

Fairly positive
Negative
Don't know

Generally speaking, do you think that the 2006 Turin Olympic Winter Games will stimulate economic recovery or not ?
Great recovery
Small recovery
Not recovery
Don't know

Which score do you think Turin municipality will deserve for the organization of the Olympic event?
A (high)
B
C
D (low)

These variables are used as "*analysis variables*". Results obtained on these items are verified by means of two other "*control variables*":
- An awareness typology (AKI);

		Knowledge Rate	
		high	*low*
Information Evaluation	*Satisfied*	***Well Informed***	***Not Interested***
	Not Satisfied	***Exacting***	***Uninformed***

- A typology (IPE) based on the degree of interest shown for other public event organized by the local administration (International Book Fair, Turin marathon …):
Fan
Participant
Indifferent
Hostile

We performed correspondence analysis to represent relationships between these items. The solution is calculated on *analysis variables*, while *control variables* are used as supplementary [3] ones.

---

[2] Original variables were constructed with five categories but in the analysis the two negative (very, fairly negative) have been aggregated.

[3] For details refer to Sas Proc Corresp.

In Fig. 1 the result of the analysis is reported. On the right side of the graph there is a refusal area: respondents disagree with the choice of hosting the Games. Further they are not proud and they think the Games will not bring advantages either to them or to the city. These respondents also think Turin city Council won't be able to manage the commitments implied for entered into the Games.

The left portion of the graph shows a more divided attitude. Moving from the lower to the upper section all the agreement indicators moves from a partial to a complete assent.

Also the *control variables* (green in Fig. 1), follow the same pattern, even if they present less extreme positions for mathematical reasons.

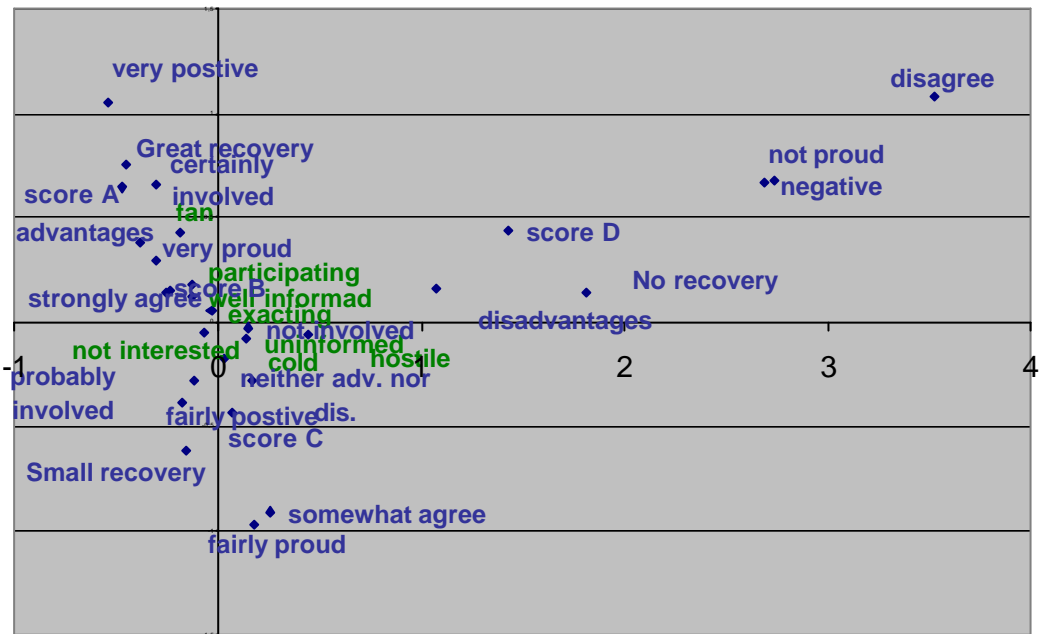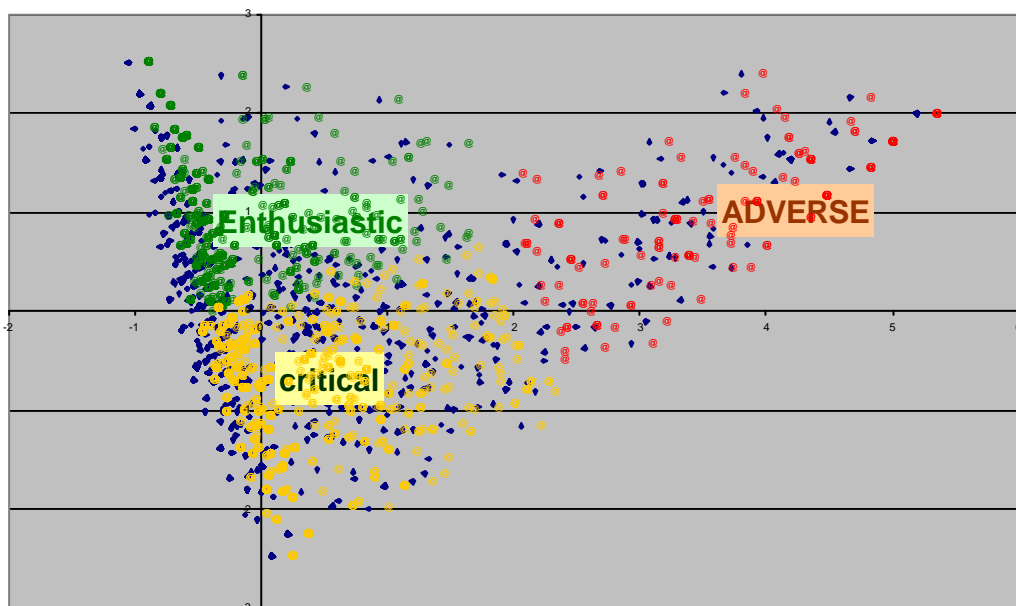Fig. 1 *Multiple Correspondece Analysis of closed questions (answers category)*

Fig. 2 *Multiple Correspondece Analysis of closed questions (subjects)*

Using the coordinates obtained by correspondence analysis as input in a cluster analysis we classified the interviewed (Fig.2). A small portion of the sample can be labelled as *Adverse* (5%). The *Enthusiastic* represent 38% of the sample whereas the majority of the interviewees are assigned to the *Critical* group. That is a group expressing favour with some criticism.

Fig. 3 *Cluster distribution*
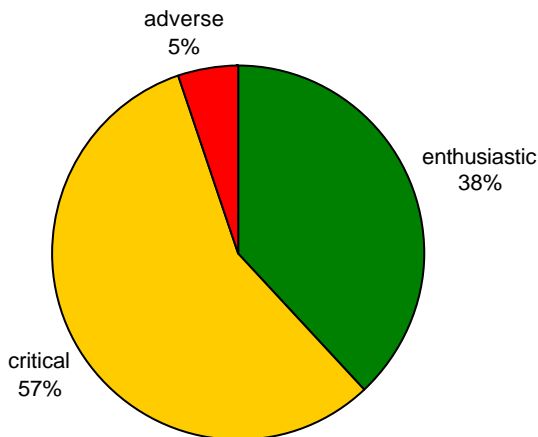
5. Step two – analysis of open questions
The first question in the questionnaire was an open one: *thinking about Turin what first comes to your mind?*
This question was asked before introducing the Games topic and so was conditioning free. For this reason in this step text mining techniques can support results obtained in the previous analysis. Resorting to an extension of correspondence analysis technique (Analysis of Lexical Table[4]), we included in the analysis answers given to the open questions as supplementary variable[5].
In this way open question have an illustrative role, helping to understand differences among the three groups previously defined.

---

[4] L.Lebart, A. Salem, *Analyse statistique des données*, 1988
[5] M. Bécue-Bertaut, J. Pragès, *Analyse conjointe de questions ouvertes et de questions fermées: methodoligie, exemple*, JADT2002: 6es Journées internationales d'Analyse des Données Textuelles (pp. 113 – 123)

After the text parsing 104 words were extracted from the start list of 202 (Fig. 4).

Fig.4 - *Most used words*
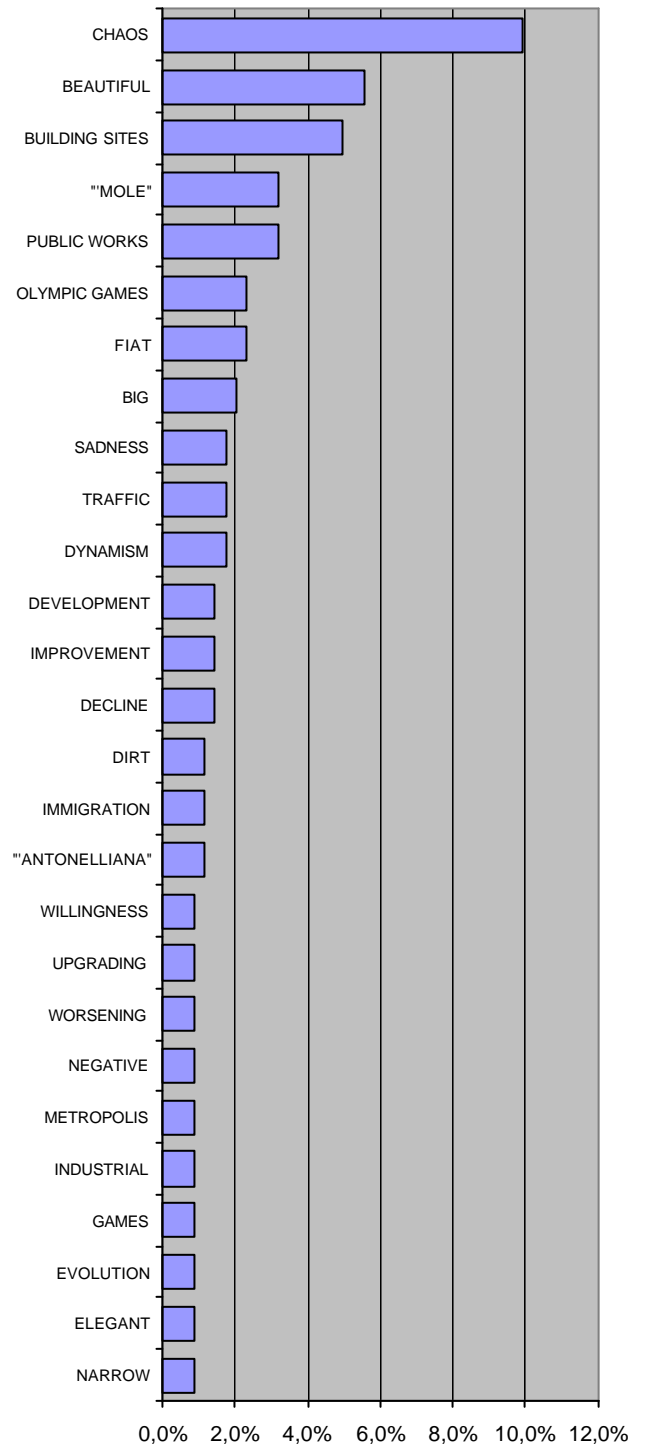
We can see different use of words in different groups, by comparing column percentages in the following table (Tab.1).

Tab. 1 - *Most used words * clusters*

Verbal form	Enthusiastic	Critical	Adverse
	Column percent		
CHAOS	8,6	13,8	8,7
BEAUTIFUL	7,5	2,6	0,0
BUILDING SITES	4,8	6,9	0,0
PUBLIC WORKS	2,7	3,5	8,7
"MOLE"	1,1	6,0	4,4
FIAT	2,7	2,6	0,0
OLYMPIC GAMES	2,7	1,7	0,0
BIG	3,7	0,0	0,0
DYNAMISM	2,1	1,7	0,0
TRAFFIC	0,0	5,2	0,0
SADNESS	2,1	0,9	0,0
DECLINE	1,1	1,7	4,4
IMPROVEMENT	1,6	1,7	0,0
DEVELOPMENT	2,1	0,9	0,0
ANTONELLIANA	1,1	0,9	4,4
IMMIGRATION	0,5	2,6	10,0
DIRT	1,1	1,7	0,0
NARROW	1,1	0,9	0,0
ELEGANT	0,5	0,9	4,4
EVOLUTION	1,1	0,9	0,0
INDUSTRIAL	1,1	0,9	0,0
METROPOLIS	1,1	0,9	0,0
NEGATIVE	1,1	0,0	4,4
WORSENING	0,0	2,6	0,0
UPGRADING	0,5	1,7	0,0
WILL	1,6	0,0	0,0
CHANGING	1,1	0,0	0,0
DISASTER	0,0	0,9	4,4
CRISIS	0,5	0,0	4,4
CRIME	0,5	0,8	0,0

It's now possible to insert these words on the graph shown in Fig. 1, in order to describe the three groups of citizens (Fig. 5).

We can see how the *Adverse* area has prevalently a negative image of the city, perhaps with an elitist shade (*elegant*). We can find words such as *disaster*, *environment* (they talk about pollution), *negative* and *crisis*. This confirms that this is the area of total refusal, in Torinese dialect *"bastian contrari"*, the ones who disagree in principle.

It follows that these citizens are maintain their convictions. Any attempt to gain the favour of these citizens (a minority group) would probably mean wasting time and resources.

At the top-left of the graph the area of complete agreement with the Olympic project coincides with a positive image of Turin. Among the words used we can find *pride, transformation, will, enjoyable* and *evolution*. It's the image of a dynamic city whose development is linked to the Olympic Games. Finally, in the lower part of

Fig. 5 *Multiple Correspondece Analysis of open question (verbal forms)*

the graph the area of *criticality*, the one on which could be useful to intervene. Citizens in this area complain about small problems of everyday life: transport, building sites, crime, immigration. Looking at the graph in Fig. 1 we see that this group of citizens is different from the *Enthusiastic* group also for the smaller degree of involvement that they perceive. This is the right target for a new information campaign and analyzing in depth the attitude of this group towards the Olympic Games it's possible to single out the more appropriate content to completely convince these citizens of the usefulness of the Olympic project.

6. Step three – analysis of target group

Critical group (target group) is composed particularly of males, of people over 51 years old, with a low level of education and resident in the Alpine area where skiing competitions will take place.

They, with the other interviewees, were asked about the possible advantages arising from hosting the Games.

In fig. 6, the percentages of those who strongly believe in those advantages, are shown for each group. *Enthusiastic* (green line) and *Critical* (yellow line) present a similar pattern, while *Adverse* (red line) shows some differences. Generally, *Critical* pattern seems to be equal to *Enthusiastic* one in the shape, but at a lower level.

In details we can observe two cases (employment and private profit) in which the yellow line is nearer to the red one than to the green one. This means that these two items are critical aspects for the target group.

Hence, as a first result, we can state that an information campaign, highlighting all the positive aspect connected with the Games and especially job opportunities, is needed.

We can now extend the same procedure to the possible problems. Fig. 7 follows the same scheme of the previous one.
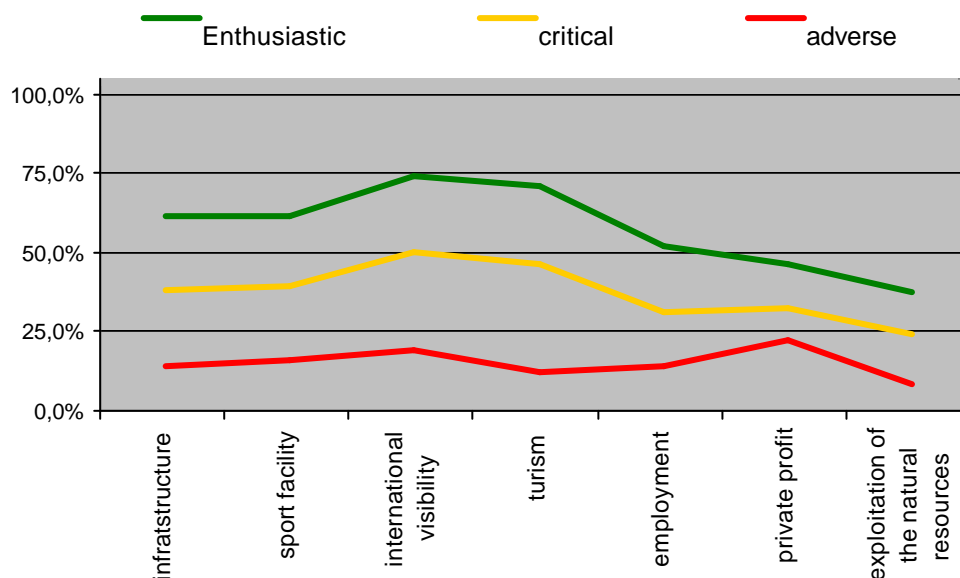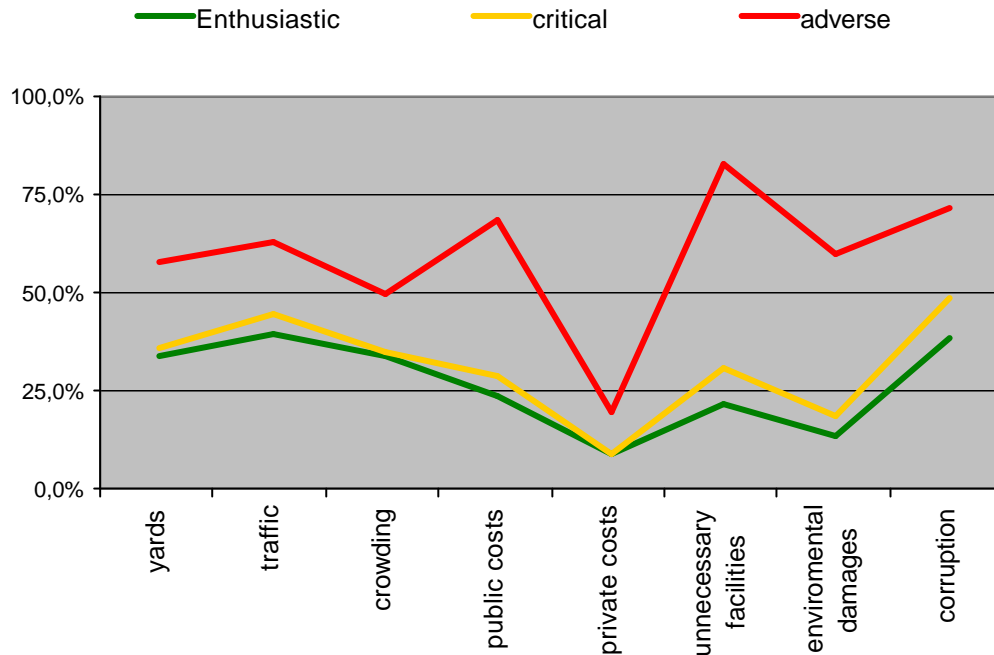
Fig. 6 *Advantages*

Fig.7 *Problems*

The red line (adverse) shows a completely different pattern from the other two, which lie very close each other. However, *Enthusiastic* and *Critical* patterns show some small differences (traffic, public costs, environmental damages) and some bigger ones (unnecessary facilities, corruption). These are the themes that could complete the information campaign addressed to the target group.

Fig. 8 *Clusters and information channels*

Finally, association between different groups and the favorite source of further information (Fig. 8) can guide local administrators in the choice of the right channel (totem, flysheets, guide).
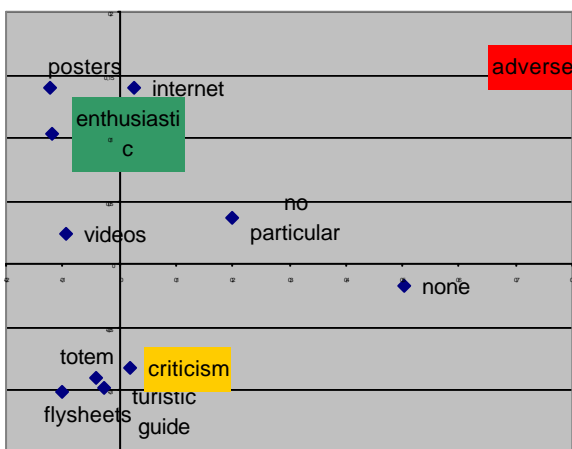
7. Conclusion

At first sight data of the Monitor of Attitude towards the Winter Olympic Games show a general citizens agreement with the choice of hosting the Games.

However, we've seen how, analyzing deeply interviewees' answers, we can succeed in singling out a particular group of citizens who are somewhat perplexed and who need more and specific information for being more enthusiastic about the project.

Text mining analysis has helped us in describing this target in a more precise way. Finally, we've been able to discover the contents the target is particular sensitive to, and the channels where they'd like to find them. These are the right ingredients for a successful information campaign.