SIS 2017
Statistics and Data Science: new challenges, new generations

28–30 June 2017
Florence (Italy)

Proceedings of the Conference of the Italian Statistical Society

edited by
Alessandra Petrucci
Rosanna Verde

FIRENZE UNIVERSITY PRESS
2017

SIS 2017. Statistics and Data Science: new challenges, new generations : 28-30 June 2017 Florence (Italy) : proceedings of the Conference of the Italian Statistical Society / edited by Alessandra Petrucci, Rosanna Verde. – Firenze : Firenze University Press, 2017.
(Proceedings e report ; 114)

http://digital.casalini.it/9788864535210

ISBN 978-88-6453-521-0 (online)

Peer Review Process
All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Committees of the individual series. The works published in the FUP catalogue are evaluated and approved by the Editorial Board of the publishing house. For a more detailed description of the refereeing process we refer to the official documents published on the website and in the online catalogue of the FUP (www.fupress.com).

Università degli Studi di Firenze
Firenze University Press
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

SOCIETÀ ITALIANA DI STATISTICA
Sede: Salita de' Crescenzi 26 - 00186 Roma
Tel +39-06-6869845 - Fax +39-06-68806742
email: sis@caspur.it web:http://www.sis-statistica.it

La Società Italiana di Statistica (SIS), fondata nel 1939, è una società scientifica eretta ad Ente morale ed inclusa tra gli Enti di particolare rilevanza scientifica. La SIS promuove lo sviluppo delle scienze statistiche e la loro applicazione in campo economico, sociale, sanitario, demografico, produttivo ed in molti altri settori di ricerca.

Organi della società:

Presidente:
- Prof.ssa Monica Pratesi, Università di Pisa

Segretario Generale:
- Prof.ssa Filomena Racioppi, Sapienza Università di Roma

Tesoriere:
- Prof.ssa Maria Felice Arezzo, Sapienza Università di Roma

Consiglieri:
- Prof. Giuseppe Arbia, Università Cattolica del Sacro Cuore
- Prof.ssa Maria Maddalena Barbieri, Università Roma Tre
- Prof.ssa Francesca Bassi, Università di Padova
- Prof. Eugenio Brentari, Università di Brescia
- Dott. Stefano Falorsi, ISTAT
- Prof. Alessio Pollice, Università di Bari
- Prof.ssa Rosanna Verde, Seconda Università di Napoli
- Prof. Daniele Vignoli, Università di Firenze

Collegio dei Revisori dei Conti:
- Prof. Francesco Campobasso, Prof. Michele Gallo, Prof. Francesco Sanna, Prof. Umberto Salinas (supplente)

SIS2017 Committees

Scientific Program Committee:
Rosanna Verde (chair), Università della Campania "Luigi Vanvitelli"
Maria Felice Arezzo, Sapienza Università di Roma
Antonino Mazzeo, Università di Napoli Federico II
Emanuele Baldacci, Eurostat
Pierpaolo Brutti, Sapienza Università di Roma
Marcello Chiodi, Università di Palermo
Corrado Crocetta, Università di Foggia
Giovanni De Luca, Università di Napoli Parthenope
Viviana Egidi, Sapienza Università di Roma
Giulio Ghellini, Università degli Studi di Siena
Ippoliti Luigi, Università di Chieti-Pescara "G. D'Annunzio"
Matteo Mazziotta, ISTAT
Lucia Paci, Università Cattolica del Sacro Cuore
Alessandra Petrucci, Università degli Studi di Firenze
Filomena Racioppi, Sapienza Università di Roma
Laura M. Sangalli, Politecnico di Milano
Bruno Scarpa, Università degli Studi di Padova
Cinzia Viroli, Università di Bologna

Local Organizing Committee:
Alessandra Petrucci (chair), Università degli Studi di Firenze
Gianni Betti, Università degli Studi di Siena
Fabrizio Cipollini, Università degli Studi di Firenze
Emanuela Dreassi, Università degli Studi di Firenze
Caterina Giusti, Università di Pisa
Leonardo Grilli, Università degli Studi di Firenze
Alessandra Mattei, Università degli Studi di Firenze
Elena Pirani, Università degli Studi di Firenze
Emilia Rocco, Università degli Studi di Firenze
Maria Cecilia Verri, Università degli Studi di Firenze

Supported by:
Università degli Studi di Firenze
Università di Pisa
Università degli Studi di Siena
ISTAT
Regione Toscana
Comune di Firenze
BITBANG srl

Index

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), CC BY 4.0, 2017 Firenze University Press

Preface

The 2017 SIS Conference aims to highlight the crucial role of the Statistics in Data Science. In this new domain of "meaning" extracted from the data, the increasing amount of produced and available data in databases, nowadays, has brought new challenges. That involves different fields of statistics, machine learning, information and computer science, optimization, pattern recognition. These afford together a considerable contribute in the analysis of "Big data", open data, relational and complex data, structured and no-structured. The interest is to collect the contributes which provide from the different domains of Statistics, in the high dimensional data quality validation, sampling extraction, dimensional reduction, pattern selection, data modelling, testing hypotheses and confirming conclusions drawn from the data. In the mention that statistics is the "grammar of data science", statistics has become a basic skill in data science: it gives right meaning to the data. Still, it isn't replaced by newer techniques from machine learning and other disciplines but it complements them. The Conference is also addressed to the new challenges of the new generations: the native digital generations, who are called to develop professional skills as "data analyst", one of the more request professionality of the 21st Century, crossing the rigid disciplinary domains of competence. In this perspective, all the traditional statistical topics are admitted with an extension to the related machine learning and computer science ones. The present volume includes the short papers of the contributions that will be presented in the 4 invited speaker sessions; in the 19 specialized sessions; in the 11 solicited sessions; in the 6 foreign societies sessions and in the 17 contributed sessions as well as, in the panel session.

Rosanna Verde
President of the Scientific Programme Committee

Alessandra Petrucci
President of the Local Organizing Committee

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), CC BY 4.0, 2017 Firenze University Press

An innovative approach for Opinion Mining : the Plutchick analysis

Un approccio innovativo per l'Opinion Mining: la Plutchick analysis

Massimiliano Giacalone, Antonio Ruoto, Davide Liga, Maria Pilato, Vito Santarcangelo

Abstract In this work we introduce an innovative approach for "Sentiment Analysis" or Opinion Mining, that is classically based on the concept that some words have positive or negative meanings. Infact, introducing the Plutchick score, it is possible to achieve an Emotional Analysis, that is a deeper analysis over the polarity. The original contribution of the paper is to present a program on Italian Emotional analysis of social networks hashtag mainly as part of "InfoSphere". For this scope we introduce AIN_EMOTION, an evolution of AIN Thesaurus, that is the first italian thesaurus for Emotional Analysis. This analysis gives a ratio of emotional hashtag on shared by social network users, can produce a behavioral trend and could be applied to any other language simply by changing the "emotional thesaurus".

Abstract *In questo lavoro si introduce un approccio innovativo per la "Sentiment Analysis" o Opinion Mining, che si basa sul concetto classico che alcuni parole assumono significati positivi o negativi. In particolare, introducendo il punteggio Plutchick, e' possibile realizzare un'analisi emotiva, cioe' un'analisi piu' approfondita sulla polarita'. Il contributo originale del lavoro e' quello di presentare un programma su Emotional in italiano utilizzando un analisi delle reti sociali hashtag principalmente come parte di InfoSphere. Per questo scopo abbiamo introdotto AIN_EMOTION, un'evoluzione del AIN Thesaurus, che e' il primo thesaurus italiano per l'analisi emozionale. Questa analisi da' un rapporto di hashtag emotiva condiviso da utenti del social network, in grado di produrre una tendenza comportamentale e potrebbe essere esteso ed applicato a qualsiasi altra lingua semplicemente cambiando il "thesaurus emozionale".*

Key words: Sentiment analysis, Instagram, Social Network, Sentiment Thesaurus, Emotion Projection

Massimiliano Giacalone, University of Naples, e-mail: massimiliano.giacalone@unina.it
Antonio Ruoto, iInformatica S.r.l.s., e-mail: marketing@iinformatica.it
Davide Liga, iInformatica S.r.l.s., e-mail: web@iinformatica.it
Maria Pilato, iInformatica S.r.l.s., e-mail: mariapilato@icloud.com
Vito Santarcangelo, University of Catania, e-mail: santarcangelo@dmi.unict.it

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online),

1 Introduction

Social networks push people to emphasize their own identity: people want to be part of certain kind of groups and, naturally, the wish of being part of some groups can be seen as the wish of not being part of many other groups. The aspect above mentioned means that social networks generate a world in which groups grow up becoming more and more separated if not openly opposed, a world in which identities are day by day clustered [8].
This sort of clustered Network Society presents some important concerns. Firstly, the more a society presents clustered identities the worse the risk of social conflicts becomes. Secondly, it seems that public opinion derived from social networks generates debates that are more emotional than reasonable. As a result, this kind of debates could be a major step towards a further consolidation of the Post-Truth politics. If the Web is the place where everybody can express his/her own opinion, and where groups develop becoming increasingly isolated or separated, social conflicts could arise[9].
The reason of such a concern could be the lack of communication between different points of view: groups are closed and self-referential, so there is less space for a true and sensible debate. This is a huge issue because the Web is luckily to be that place where public opinion will be generated in the future. Moreover, the more groups become isolated and self-referential the more debates become emotional. In this regard, it should also be noticed that overload information can exercise a significant influence on individual choices and groups. It seems that this kind of information reduces the awareness of ideas, pushing also people to rely on groups and collective identities or even swapping from one group to another. If public opinion becomes emotional and volatile, political powers will try to adjust according to this lack of awareness. For example, that could be one of the reasons why western societies are facing the so-called populisms [10].
As can be seen, there is a good reason to believe that it is vital to develop an in-depth understanding of how human emotions and Network Society are related. As said, public opinion is becoming more emotional and volatile, in addition to this, societies are more clustered and potentially conflictual, and all these aspects seem to be politically relevant. In this scenario, social networks are luckily to become increasingly important in the generation of the common sentiment: they are the place where individual choices can be influenced and public opinion can be interpreted [11].

2 The social network in the infosphere semantic space

The following research aims to outline a survey methodology that can map the emotional level of the great Italian conversation on the network as part of the broader context of the infosphere. Specifically, it will be given an application of this methodology grafted on the social network "Instagram" with the aim of understanding with

regard to the Italian language users: 1) In what direction it propagates the generated emotional strength; 2)What are the main conveyed emotional projections; 3) What are the 10 most significant emotional biases conveyed. The term in the information

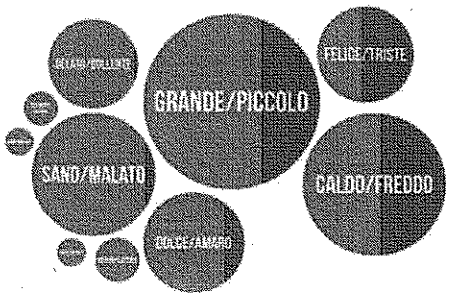

Fig. 1 The 10 main polarizations

infosphere philosophy means the totality of the information space. The infosphere is "the semantic space constituted by the totality of the documents, the agents and their operations", where "documents" means any data, information and knowledge, codified and implemented in any semiotic format, "agents" are any system able to interact with an independent document (such as a person, organization or web software robots) and the term "operations" includes any kind of action, interaction and transformation that can be done by an agent and which can be presented in a document. It is an environment in which the organisms are formed as interconnected cells [1].

It is immediately evident, within the perspective of contemporary Network Society, as the current dynamics of the Great Conversation network is configured as an essentially structural part of the infosphere. The spread of the Internet, mobile communications and digital media, along with a wide range of social software tools are driving the development of interactive and horizontal communication networks that connect, at any time, local and global. The communication system of industrial society revolved around the mass media, characterized by the mass distribution of a one-way message one-to-many, one to many. The communication foundation of the network society is the global system of horizontal communication networks, which include the multimodal exchange of many-to-many interactive messages, or many to many, synchronous and asynchronous. If we consider the public sphere as the space in which form public opinion, the analysis of the dynamics at play within the great conversation on the network, affected by the events and the agenda setting issues, can be useful to read the major changes taking place, producing predictive projections on generation of common sense and construction mechanisms of collective and individual imaginary [2].

However, it must take note that public opinion structured along the great network conversation is giving an opinion I can essentially emotional. The sharing of infor-

mation overload and Emotional effects are transforming the public debate in a more emotional than rational debate [3, 4].

3 Methodology

In the Network Society then the only effective message is an emotional message, which starts from something emotionally powerful. The amplification of the emotional sphere becomes the basis on which the circulation of information in the sphere of mass self communication. A movement that runs along the paths laid out by the emotional contagion mechanism. It is the confirmation of the claims of almost three centuries by philosopher David Hume is the reason to be the slave of the emotions, and not viceversa [5].
If then you become aware that in this context at the base of the social movements decisions there are processes that involve the transformation of the social emotions, emotional intelligence operation allows you to frame and define the extent of the emotional dynamics in place able to compete in processes of creation of common sense and influence the construction of collective and individual imaginary [7].
The mapping of the emotions conveyed in the large network conversation is carried out in this work through the methodology OSINT, proceeding to an activity of gathering information by consulting publicly available sources. Specifically, they are classified on the basis of an emotional score major adjectives Italian language as used by users of Instagram and hashtag products from 10/10/2010 (the year of the launch platform in Italy) to 10/01/2016 with a recurrence greater than or equal to 100,000. Emotional score attributed to each hashtag is based sull'AIN thesaurus, among the most comprehensive thesaurus for analysis sentiments of the Italian language, which attaches to the main adjectives of the Italian language a positive or negative emotional polarity following a scoring system based on the following ladder: + 2 (very positive), + 1.5, + 1, 0.5;
0 (neutral);
- 0.5, - 1, - 1.5, - 2 (very negative).
In addition to clarifying the emotional polarity of each hashtag analyzed, these have been classified through the emotional scale proposed by the American psychologist Robert Plutchik [6].

AIN_EMOTION is the first thesaurus of "emotional" type for the Italian language. The AIN Thesaurus which is the thesaurus for opinion mining (positive, negative, neutral) the Italian language has been enriched the speech "EMOTION" categorizing each adjective with the emotional scale Plutchick. The following is an example, where for each adjective was associated one or more clusters of Plutchick. For each word (abandoned, lowered, down, propertied, combined, plentiful, tanned, abominable, endearing) there is a relative cluster associated (pain, sadness, acceptance, sadness, serenity, interest, admiration, disgust).
Thanks to information of the emotional scale it is possible to make an opinion upper

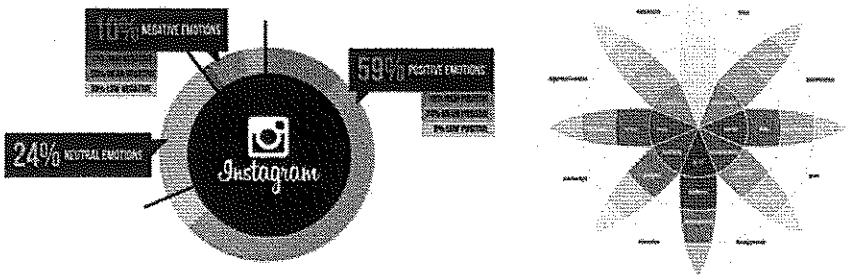

Fig. 2 Sentiment Analysis on Instagram and Plutchick Wheel of Emotions

level mining,not limited to one polarity but adding semantic information of particular importance, in order to carry out an investigation more focused on the emotional state of the author. In fact, this approach introduces an "emotional" analysis, thanks to the use of the granularity of the Plutchick score. The traditional sentiment analysis considers only 3 classes (negative,neutral,positive) with a possible weight for a better analysis. In the Plutchick approach we have 32 possible classes, 8 with 3 different degrees (e.g. admiration,trust, acceptance) and 8 intermediate (love, submission, awe, disapproval, remorse, contemp, aggressiveness, optimism) as shown in Fig.2.

4 Discussion of results and conclusions

From the surveys conducted it is summarized possible to state that, with regard to users of Italian language, the social network instagram: 1) conveys average positive emotional biases: among analyzed hashtag, those used by users most often have an emotional rather than a positive polarity; very few, in parallel, are in effect the hashtag with an average negative emotional polarity; 2) It constitutes an environment in which the most represented social emotional classes are those relating to dell' "expectation sphere" and the "pain". Almost nonexistent turn out to be the emotions that belong to the emotional classes of "anger" and "disgust".

Similar results should be read in light of an observation: the form of the content and architecture of social environments are more important than the content: from communication processes switching to narratological environments. And the stories conveyed by Instagram users, because of the nature of the medium, are on average functional to one goal: the digital packaging of the self. In this sense the self-production of content carried converges towards selfmarketing with the aim, more or less conscious and said, to show up on the market relations in a positive light as much as possible. A trend that seems to respond to a need for performance in a society where it seems that you have to be seen to exist. No wonder then that the emotion conveyed through the hashtag analyzed are crushed on emotional biases of almost exclusively positive type. It is a very precise seduction strategy of nego-

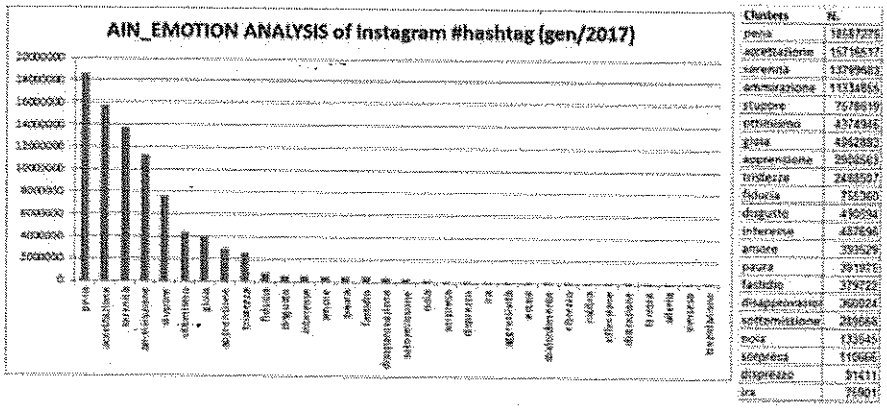

Fig. 3 Emotions extracted by M. Di Lecce tool

tiating type: convey positive emotions to receive positive in return. In this sense then on Instagram it seems to be banned any type of really critical about the real comparison. Evidence that is even more overwhelming if you take into account a relational dynamic imprinted on a omofiliaco logic. In this sense the emotional dynamics conveyed through Instagram eventually attributed to likeability, the principle of "pleasure", the only true God to be served in the development of a strategy of attention conveyed through its social narratives, much of the power creation of common sense.

References

1. Floridi L., (2012), La rivoluzione dell'informazione, Codice Edizioni, Torino.
2. Castells M., (2009), Comunicazione e potere, Universita' Bocconi Editore, Milano.
3. Lovink G.,(2012), Ossessioni collettive. Critica dei social media, Universit Bocconi Editore, Milano.
4. Carr N.,(2011), Internet ci rende stupidi?, Raffaello Cortina Editore, Milano.
5. Westen D., (2008), La mente politica, il Saggiatore, Milano.
6. Plutchik R.,(2002), Emotions and Life: Perspectives from Psychology, Biology, and Evolution, Washington, DC: American Psychological Association.
7. Santarcangelo, V., Oddo, G., Pilato, M., Valenti, F., Fornaro, C., (2015). Social Opinion Mining: an approach for Italian language, SNAMS2015 at FiCloud2015.
8. Santarcangelo, V., Pilato, M. et al. (2015). An opinion mining application on OSINT forthe reputation analysis of public administrations, Choice and preference analysis for quality improvement and seminar on experimentation, Bari.
9. Iorio, E., Ruoto, A., (2015). Nessun Tempo. Nessun Luogo. La comunicazione pubblica italiana all'epoca delle Reti.
10. Pilato, M., Santarcangelo, G., Santarcangelo, V., Oddo, G., (2015). AIN Thesaurus, RCE MULTIMEDIA.
11. Giacalone M. (2009) Manuale di Statistica Giudiziaria "Bel Ami edizioni" Roma.
12. Ruoto A., Santarcangelo V., Liga D., Oddo G., Giacalone M., Iorio E. (2017). The sentiment of the infosphere: a sentiment analysis approach for the big conversation on the net. In: (a cura di): Lauro C. Amaturo E. Grassia M.G. Aragona B. Marino M. (Eds.), STUDIES IN CLASSIFICATION, DATA ANALYSIS, AND KNOWLEDGE ORGANIZATION, Springer.