Social Media Geographic Information Visual Analytics

Junia Borges[a], Ana Clara Moura[a], Priscila de Paula[a] and Pedro Casagrande[b]

[a] Urban Planning Department, Federal University of Minas Gerais, Belo Horizonte, Brazil (juniaborges@yahoo.com.br, anaclara@ufmg.br, priscilalpaula@msn.com)
[b] Geography Department, Federal University of Minas Gerais, Belo Horizonte, Brazil (pedrobcasagrande@gmail.com)

Key-words: Volunteer Geographic Information, Social Media Geographic Information, Image Analysis, Passive Volunteer Geographic Information, Instagram.

Introduction

The use of Social Media Geographic Information in a planning process would improve our knowledge on urban and landscape development and support decision-making. This paper makes use of an environmental disaster that has happened recently in one of the most important social economic areas in Brazil to understand how this type of information could be used as a systematized planning input. Researchers seeks to understand if it is possible to establish a connection between SMGI posts and the level of affection to shared content. In order to do that, we tested a passive SMGI analysis. Our best results so far relied on a proposal for image analysis from Instagram posts.

Analytics gathered from Social Media Geographic Information (SMGI) improve comprehension of community values (Borges, Jankowski and Davis Junior, 2015b). There are several information attributes made available by ordinary citizens within social network environments such as location, texts, videos, preferences, images and audio. The development of a systematic methodology to analyze information posted by ordinary people would contribute to landscape and urban planning processes. This paper presents a new methodological direction to investigate image collection from Social Media.

In Volunteer Geographic Information (VGI), citizens act as sensors. Its development leads to innovative models using social networks and is being denominated Social Media Geographic Information. Campagna (2015a) understands it as a deviation from a traditional vision of VGI and states that it may be used for both leisure and professional reasons. Ideally, it would allow integration and sharing of the resulting information flow.

In spatial planning it is highly recommended that citizens be heard. Acknowledgment and comprehension of the most important value variables for a group of people are keys to landscape and urban planning. After a serious environmental disaster happened on November 5, 2015 in the Quadrilátero Ferrífero (Iron Quadrilateral) region in Minas Gerais, Brazil, a case study using SMGI from an passive perspective was developed. The strategy of investigating passive SMGI was the analysis of Instagram images.

The purpose of this research is contributing towards identification of collective values, aiding local image reconstruction as well as subsidizing a landscape recovery plan in a new hazard area. The hypothesis pursued is determining whether it is possible or not to establish a connection between SMGI posts and the level of affection to shared content.

Methodology

A "passive VGI" analysis from Instagram messages was performed using LETICIA API11, a courtesy of LABIC (Laboratório de Imagem e Cybercultura - Laboratory of Image and Cyberculture of Federal University of Espirito Santo, Brasil). Messages since the day/time of the incident were downloaded aiming at analyzing values, behaviors and profiles of the posts. The expectation was to identify the *genius loci*(the essence of the place and its symbols) through the collective description and investigation of images, videos, text and geolocation of posts.
According to Borges, Jankowski and Davis Junior (2015a) the hashtag (#) is the main way to attach a category or a subject to a message in a social media environment. The authors point out that a semantic analysis should follow the potential words used by the group of people that is the research agenda's focus. This could also help by indicating the profile of people related to the certain topic. Hashtags were selected in two ways: firstly, by researchers' intuition #mariana #riodoce and secondly using a platform called Tagboard (available at tagboard.com), a real time search engine that favors the investigation of a hashtag, it was found a third tag: #sosriodoce. After a locational analysis of posts, from the three tags, the last was defined as most appropriate as they have a better rate of onsite posts. The download was done from November 5, 2015 to November 27, 2015.
Posts were collected and selected within a 10km distance from the affected area (Rio Doce River) to perform an image analysis was proposed to separate drawing, landscape and everyday life images. We imported photos into ArcGIS 10.2 and sliced the photos into 5 different colors. The reason why we choose five is based on studies of mental maps on semantic applied to cartography, that states the human capacity to understand and read the differences (Bertin, 1977).
A statistic analysis to show the standard deviation overall of polygons' areas was next step. Finally image classes were defined regarding its standard deviation pattern in order to separate them into three image pattern: everyday life, landscape and meme (broadly spread multimedia content via internet).

Results and discussion

The everyday life (resulted in many small polygons in the picture (Figure 1 C) and the standard deviation among the shapes' areas is low. The landscape photo segmentation resulted in large shapes, but also in small shapes in some parts, what results in medium standard deviation of polygon's' areas. The drawing memes resulted in very large areas in contrast with very small areas in the segmented representation, what presents a high standard deviation among the shapes' areas (Figure 1 I).

[11] https://github.com/andreibastos [Online: accessed March 20, 2015].

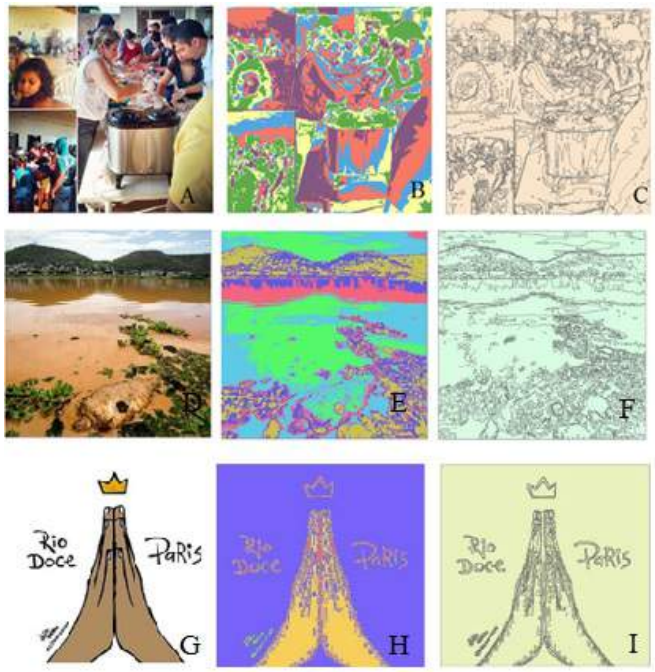
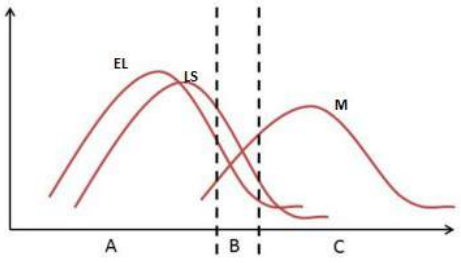


Fig. 1. Image study using ArcGIS 10.2. A: photo of everyday life. D: photo from the landscape. G: drawing meme. B/E/H: sliced into 5 classes. C/F/I: polygonal study.

The images selected were automatically identified (everyday life, landscape and meme) and confirmed by a manual visual classification. Then the statistical analysis was performed to confirm the results.

Images of classification Color Standard Deviation

Breaks : A - B - C
A → <= 730
B → 730 < B < 1010
C → >= 1010



Break	Total images	Images meme	Images landscape	Images everyday life
A	114	19	42	53
B	20	8	7	5
C	55	36	12	7

Notes on break A: p + vc (A1 + A2)
A2 → <= 400
A1 → 400 < A1 < 730

Landscape and everyday life is the standard deviation very close.

The classification method is valid, whereas the percentage of error in each interval is 10%.

Fig. 2. Statistic results using Minitab (www.minitab.com, accessed in april 2016) standard deviation performance of polygons' areas, where EL is everyday life, LS is landscape and M is meme.

A, B and C represent respectively 'everyday life', 'landscape' and 'meme' standard deviation of polygon's areas (Figure 2). The statistical results (show) fragmented images of everyday life with many polygons. The standard deviation found amongst everyday life's areas has a low reference. The landscape results into medium areas in the picture when compared to everyday life and image memes. The drawings or memes result into big areas with high standard deviation.

The quantitative analysis of the procedure applied at #sosriodoce dataset sample shows that the results confirm our first manual analysis with an overall confusion of 10%. Landscape was confirmed as the most challenging class as it has the highest mistake rate. Landscape images essentially mixes results with Everyday Life. This confirmed the insight that a second filter should be proposed to better separate Everyday Life and Landscape classes of image posts.

Conclusions

Passive SMGI is useful for understanding citizens' values and increase knowledge for spatial planning. It represents a rich universe of information for research purposes as it needs less motivation efforts to get people involved as well as less investment in publicity.
A connection between SMGI passive posts and the level of affection to shared content shall be further investigated using a broader sample. The right choice of a hashtag (a dataset primary characteristic) can determine the success of genius loci catching. An analysis of number of "likes" could also be carried out by image categories and correlating them to location.
Further investigations into image classification methods should be done. Image classification was very efficient to separate drawings from the other two categories of everyday life and landscape. Confusion between everyday life and landscape is explained by the fact that it is commonplace to have people in landscape pictures. A second filter and/or image classification should be considered. We can suggest color pattern analysis as a possible path to follow.
The genius loci (the essence of the place, and its symbols) can be further investigated through the collective description of images, videos, text and geolocation of posts. The automatic process of the image classification can benefit the exploration into the values and behaviors of citizens involved in SMGI passive posts.
It is clear, especially in Brazil, that social media is used as a mechanism to manifest social and cultural issues. Memes with impact images or drawings with a critique were broadly shared. They represent the collective thought. Quantitative identification (using same area's standard deviation value) and qualitative investigation of memes repetition should lead to good insights about collective values and thoughts. However, in the directly affected area a majority of posts refer to reality, meaning that it is also used as a mechanism to register occurrences.
In passive analytics, the aim is to capture information on society's values and behaviors. That involves other issues that should be further analyzed: Does the sample correspond to reality? Does the profile of posters refer to local people? Can this investigation be performed in different geographic areas? Is this methodological proposal efficient in identifying collective behaviors? Or is it only a registration mechanism? Or maybe it is both?.

References

Bertin, Jacques. 1986. A neográfica e o tratamento gráfico da informação, Curitiba: Editora da Universidade do Paraná.

Borges, J.; Jankowski, P.; Davis Junior, C. 2015. "Crowdsourced information from Tweets during the WorldCup in Brazil: A theme search." Paper presented at Changing Cities II, Spatial, Design, Landscape & Socio-economic Dimensions.

Borges, J.; Jankowski, P.; Davis Junior, C. 2015. "Crowdsourcing for Geodesign: Opportunities and Challenges for Stakeholder Input in Urban Planning." In Cartography - Maps Connecting the World, edited by Claudia Robbi Sluter, Carla Bernadete Madureira Cruz, Paulo Márcio Leal de Menezes, 361-373. Switzerland: Springer.

CAMPAGNA, M. 2016. "Social Media Geographic Information: Why is special when it goes spatial?" Energic Book. Ubiquity Press.

Websites

www.minitab.com

https://github.com/andreibastos