Semantic and Geospatial Mapping of Instagram Images in Saint-Petersburg

Yuri Rykov, Oleg Nagornyy, Olessia Koltsova
National Research University Higher School of Economics
St.Petersburg, Russia
{yrykov, onagorny, ekoltsova}@hse.ru

Herbert Natta
University of Rome Tor Vergata,
Rome, Italy
herbert.natta@gmail.com

Alexander Kremenets
makeomatic
Moscow, Russia
kremenets0207@gmail.com

Lev Manovich
City University of New York
New York, USA
manovich.lev@gmail.com

Damiano Cerrone
Spatial Intelligence Unit
Tallinn, Estonia
damianocerrone@gmail.com

Damon Crockett
Software Studies Initiative
La Jolla, USA
damoncrockett@gmail.com

The availability of large urban social media data creates new opportunities for studying cities. In our paper we propose a new direction for this research: a joint analysis of geolocations of shared images and their content as determined by computer vision. To test our ideas, we use a dataset of 47,410 Instagram images shared in the city of St.Petersburg over one year. We show how a combination of semantic clustering, image recognition and geospatial analysis can detect important patterns related to both how people use a city and how they represent in social media.

I. Introduction

New digital urban studies is the prospective interdisciplinary research area that combines issues and methods of urban sociology, computer science, digital humanities, linguistics, design and architecture to retrieve knowledge about everyday life and social organization of cities from diverse data sources [1]. Digital urban studies may be also viewed as a part of cultural analytics approach developed in [2], [3] aimed at seeking connections between seemingly unrelated aspects of urban life. As a new sub-discipline, it demands new research designs that combine methods from the aforementioned fields.

An important task of digital urban studies is to extract meanings that visitors assign to different urban areas, and one of the ways to do it is to find relations between the content of digital traces that users leave online and the geographical location of those traces. In this paper, we seek to examine whether it is possible to solve such task by finding connection between content and geotags of Instagram images shared in city space [4]. We ask whether such images can be clustered into meaningful categories reflecting human experience and whether this experience can be related to certain urban areas in a meaningful way. We believe that such results can be used to study urban segregation, as well as to rate city areas in terms of their consumer/tourist attractiveness or cultural/entertainment development.

II. Related work

Works in digital urban studies have exploited user check-in behavior to investigate organization of urban space [5], as well as other geolocated user activity to map urban social media inequality [6]. However, image analysis is virtually absent from urban studies. Boy et al [7] compared spatial patterns and divisions of user communities defined through interaction networks based on Intagram data from Amsterdam and Copenhagen. However, they did not investigate any content or visual properties of images which is the most essential feature of Instagram, and our research seeks to cover this gap.

III. Data and methods

A. Instagram dataset description

We use Instagram dataset of 47,410 items from Saint-Petersburg city, Russia (excluding the district east to a conventionally selected longitude, for technical reasons). Each item contains an image, a time stamp, geographical coordinates, user's ID and user-generated #hashtags. This dataset was collected with Instagram API during one year period from July 1 2014 to June 30 2015.

These images were processed with Google Cloud Vision API service. As a result these images received 3,882 artificial tags describing recognized entities, one or more tag per image. Each tag has a probability value reflecting a veracity of entity recognition. In this pilot research all Google-defined tags were used without weights, although in future it would be useful to include them.

B. Semantic network analysis

To extract semantic domains, to which images may be assigned, we represent our data as a network of tag co-occurrence with IFRIS Cortext Manager online software (http://managerv2.cortext.net/). Multiplicity of Google-defined tags per image is essential for this task. Our graph is undirected and weighted. Each vertex represents a particular Google-defined tag. Each arc represents a measure of

ISBN 978-952-68397-8-3 (PDF)

similarity based on normalized co-occurrence of a particular pair of tags in a set of tags ascribed to the same image. The more frequently two tags co-occur, the more similar they are. To normalize co-occurrence data and derive a similarity score, we applied distributional normalization procedure implemented in Cortext software [8], [9]. The resulting distributional similarity of tags is equivalent to distributional similarity of words defined in [10] as “the extent to which they can be inter-substituted without changing the plausibility of the sentence”.

To cluster the graph of tags, Louvain clustering algorithm was applied [11] as the fastest and still yielding high-quality results. To measure semantic proximity between each pair of clusters correspondence analysis was performed. Correspondence analysis is based on Chi-squared similarity measure and is usually presented on the log scale implemented in Cortext software.

C. Geospatial and temporal distribution analysis

Next, each image was assigned to one or several semantic clusters with Cortext software. All images appearing in a particular cluster were plotted on the geographical map of St.Petersburg according to geospatial coordinates of those images with QGIS software. Thus each semantic cluster was represented in the city space. Finally, we compared those representations to determine whether any topics have similar geographical distributions. To do so, we computed our own version of spatial correlation coefficients between density-based heat-maps of each pair of clusters. Our spatial correlation coefficient is a normalized root-mean-square difference between heat-maps, calculated based on the histogram of the corresponding images difference.

IV. RESULTS

A. Semantic (Google-defined tag) network mapping

The resulting semantic network comprised 479 most frequent Google-defined tags belonging to the single connected component and 2,942 links among them. Louvain clustering algorithm identified 15 clusters (Fig. 1). We labeled them manually (Table I) as the quality of automatic labeling we had tried was insufficient. While the network density is low (0.029), its modularity is very high (0.859) indicating a pronounced cluster structure.

Correspondence analysis of cluster label co-occurrence shows that just few pairs of clusters are thematically intersected (Fig. 2) which again confirmed a well-defined cluster structure. The most thematically intersected pairs of clusters are “facade & palace” - “sunrise & sea” and “dish & food” - “drink”.

B. Geospatial mapping of semantic clusters

Different sematic clusters yielded different spatial and time distributions, which is potentially useful for understanding semantic organization of urban space. For instance, images from “animals” cluster are relatively evenly distributed over space (fig.3) and time of the year (not shown) which suggests “animals” to be a constant background topic of urban life showing no inequality. Also, this cluster’s daily distribution yeilds random peaks.

Contrary to this, images of “sunrise & sea” cluster appear mostly near water zones including rivers, channels and seaside with greater density in the city center (due to Niva river, channels and bridges downtown location, see Fig.4). Daytime distribution yields two clear peaks: dawn and sunset (Fig.6); monthly distribution shows a seasonal effect: much more images were taken in summer time (Fig.8). It is plausible that with a more fine-grained cluster analysis temporary event- or problem-related clusters can be identified that, moreover, can concentrate in certain areas and indicate social tensions.

TABLE I. CLUSTER LABELS AND IMAGES COUNT

Cluster label	Number of images in cluster	Relative frequency
sculpture & statue	1327	3,8%
clothing & fashion	3614	10,5%
dish & food	3590	10,4%
flowers & plants	2516	7,3%
hairstyle	4920	14,2%
art & preformance	2247	6,5%
automobile	1730	5,0%
portrait	4545	13,2%
animals	1595	4,6%
drink	1219	3,5%
christmas decorations	797	2,3%
sunrise & sea	2789	8,1%
power training	2607	7,5%
facade & palace	3697	10,7%
document & mobile device	5092	14,7%

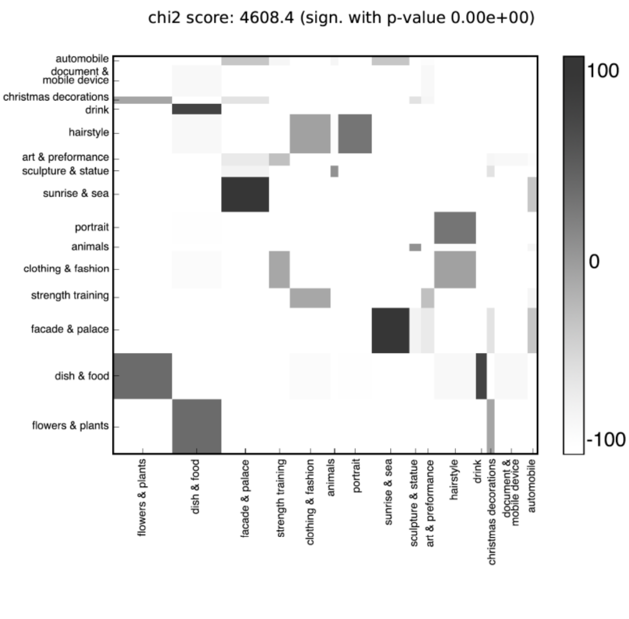

Fig. 2. Semantic clusters intersection/proximity

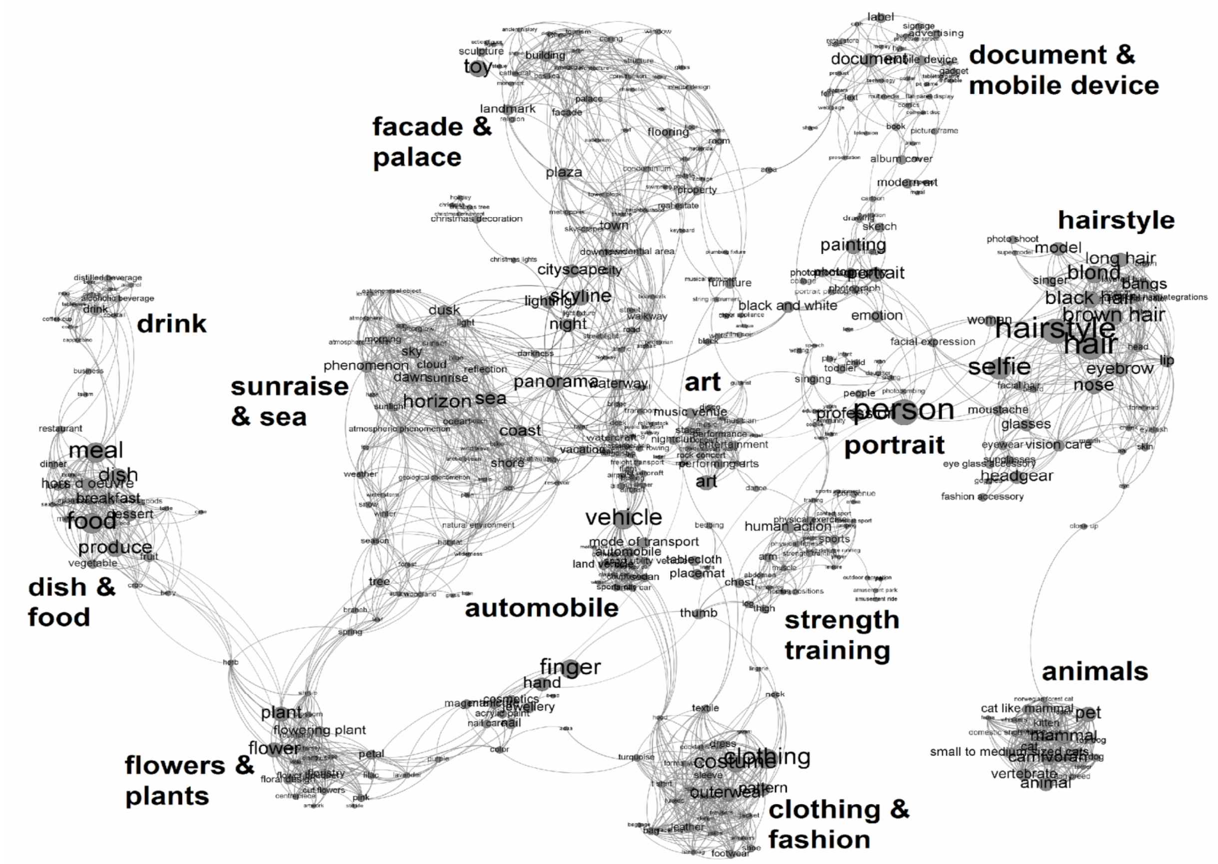

Fig. 1. Semantic (Google-defined tag) network with cluster labels

Fig. 3. “Animals” cluster posts in St.Petersburg city

Fig. 4. “Sunrise & sea” cluster posts in St.Petersburg city

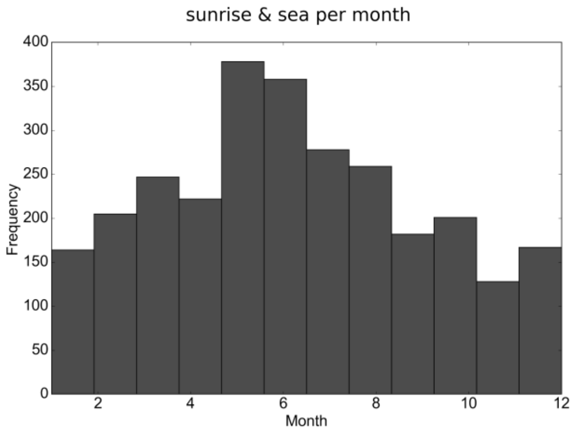

Fig. 8. “Sunrise & sea” cluster posts by months

Comparison of semantic clusters by their geographical proximity has yielded mixed results (Fig. 9). The most similar pair of clusters is “hairstyle” / “animals”. Clusters of the pair are also similar to many other clusters because they are relatively evenly distributed in space: they are geographically independent topics. Many other less similar pairs demands additional data for interpretation. We found only one case of relatively strong proximity ready for meaningful interpretation: ”facade & palace” and “art & performance” clusters are close to each other because cultural events mostly take place in or near ‘architecturally rich’ buildings with attractive facades. On the whole, semantic and geographical similarities of clusters do not highly correlate, as defined by Spearman correlation (r=0.29) which confirms the demand for further information to interpret both types of similarities. Moreover, even this level of correlation may be overestimated since similar semantic clusters tend to share many images and hence their locations.

	clo	per	flo	chr	por	dri	scu	hai	dis	fac	aut	sun	doc	arm	car
carnivoran & mammal	0.21	0.4	0.36	0.38	0.3	0.42	0.49	0.65	0.41	0.37	0.46	0.4	0.49	0.23	1.0
arm & strength training	0.0	0.24	0.29	0.05	0.08	0.19	0.21	0.34	0.03	0.17	0.26	0.21	0.09	1.0	0.23
document & mobile device	0.26	0.28	0.26	0.33	0.2	0.3	0.48	0.51	0.34	0.27	0.38	0.28	1.0	0.09	0.49
sunrise & sea	0.17	0.48	0.36	0.24	0.33	0.33	0.37	0.46	0.19	0.26	0.35	1.0	0.28	0.21	0.4
automobile & compact car	0.28	0.37	0.44	0.27	0.19	0.33	0.46	0.58	0.16	0.31	1.0	0.35	0.38	0.26	0.46
facade & palace	0.11	0.24	0.3	0.14	0.12	0.23	0.44	0.41	0.15	1.0	0.31	0.26	0.27	0.17	0.37
dish & hors d oeuvre	0.04	0.17	0.08	0.41	0.29	0.32	0.26	0.4	1.0	0.15	0.16	0.19	0.34	0.03	0.41
hair & hairstyle	0.27	0.47	0.51	0.47	0.38	0.54	0.57	1.0	0.4	0.41	0.58	0.46	0.51	0.34	0.65
sculpture & statue	0.25	0.33	0.39	0.25	0.21	0.33	1.0	0.57	0.26	0.44	0.46	0.37	0.48	0.21	0.49
drink & pint us	0.06	0.28	0.31	0.42	0.31	1.0	0.33	0.54	0.32	0.23	0.33	0.33	0.3	0.19	0.42
portrait & photography	0.04	0.38	0.19	0.35	1.0	0.31	0.21	0.38	0.29	0.12	0.19	0.33	0.2	0.08	0.3
christmas decoration & christmas tree	0.1	0.27	0.17	1.0	0.35	0.42	0.25	0.47	0.41	0.14	0.27	0.24	0.33	0.05	0.38
flower & cut flowers	0.18	0.37	1.0	0.17	0.19	0.31	0.39	0.51	0.08	0.3	0.44	0.36	0.26	0.29	0.36
performance & concert	0.18	1.0	0.37	0.27	0.38	0.28	0.33	0.47	0.17	0.24	0.37	0.48	0.28	0.24	0.4
clothing & outerwear	1.0	0.18	0.18	0.1	0.04	0.06	0.25	0.27	0.04	0.11	0.28	0.17	0.26	0.0	0.21

Fig. 9. Geospatial similarity (proximity) of semantic image clusters

V. Conclusion

In this work we have applied automatic image recognition (computer vision) together with geospatial analysis to studying spatial organization of semantic domains generated by social media users in an urban space. To our knowledge, such combination of methods has not been applied to digital urban studies before. We show that such research design may yield meaningful results, however, more fine-grain clustering, richer user data and additional information are needed for reaching non-trivial conclusions. In the future research it is desirable to use more images retrieving them from multiple sources and to enrich geolocation with additional social information – for instance, to divide urban space into real estate price zones or in some other meaningful way.

Acknowledgments

The study was implemented in the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) in 2016. This research was started at Summer Lab "Digital Traces I: Meta-Morphologies of St. Petersburg", organized by the STS center at the European University at St.Petersburg by Diana Kurkovsky West & Vincent Lepinay, and directed by Lev Manovich & Damiano Cerrone. The dataset was prepared by SPIN Unit (Damiano Cerrone) with contributions from Software Studies Initiative (Damon Crockett). Special gratitude to Jean-Philippe Cointet for his guidance and software aid.

References

[1] A.A. Salah, L. Manovich, A.A. Salah, and J. Chow, “Combining cultural analytics and networks analysis: Studying a social network site with user-generated content”, *Journal of Broadcasting & Electronic Media*, 2013, vol.57 (3), pp. 409-426.

[2] L. Manovich, “Cultural analytics: visualising cultural patterns in the era of “more media”, *Domus,* 2009. Web: http://manovich.net/content/04-projects/063-cultural-analytics-visualizing-cultural-patterns/60_article_2009.pdf

[3] L. Manovich, “The science of culture? Social computing, digital humanities and cultural analytics”, *Manovich.net,* 2015, Web: http://manovich.net/content/04-projects/086-cultural-analytics-social-computing/cultural_analytics_article_final.pdf

[4] J. Thatcher, “Big Data, Big Questions | Living on Fumes: Digital Footprints, Data Fumes, and the Limitations of Spatial Big Data”, *International Journal of Communication*, 2014, vol.8, p. 19.

[5] J. Cranshaw, R. Schwartz, J.I. Hong, and N. Sadeh, “The livehoods project: Utilizing social media to understand the dynamics of a city”. *In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, 2012, p. 58-65.

[6] A. Indaco, and L. Manovich, “Urban social media inequality: definition, measurements, and application”, *arXiv:1607.01845 [computer science]*. 2016. Web: http://arxiv.org/abs/1607.01845

[7] J.D. Boy, and J. Uitermark, “How to Study the City on Instagram”, *PLOS ONE*, 2016, vol. 11, No. 6, p. e0158161. http://doi.org/10.1371/journal.pone.0158161

[8] A. Rule, J.-P. Cointet, and P.S. Bearman, “Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790–2014”. *Proceedings of the National Academy of Sciences*, 2015, vol.112, No. 35, pp. 10837–10844. http://doi.org/10.1073/pnas.1512221112

[9] J. Weeds, and D. Weir, “Co-occurrence retrieval: A flexible framework for lexical distributional similarity”, *Computational Linguistics*, 2005, vol.31, No. 4, pp. 439–475.

[10] K.W. Church, W. Gale, P. Hanks, D. Hindle, and R. Moon, “Lexical substitutability”, *In Computational Approaches to the Lexicon (Ed by B. T. S. Atkins and A. Zampolli)*, Oxford University Press, Oxford, 1994, pp. 153–177.

[11] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks”, *Journal of Statistical Mechanics: Theory and Experiment*, 2008, vol.10, p.10008, http://doi.org/10.1088/1742-5468/2008/10/P10008.