Approach to Automated Storytelling Using Cluster Polygons

Raunak Sarbajna
Department of Computer Science
Lamar University
Beaumont, Texas, USA
Email: rsarbajna@lamar.edu

Sujing Wang
Department of Computer Science
Lamar University
Beaumont, Texas, USA
Email: sujing.wang@lamar.edu

I. Methodology

A. Change Analysis Predicates

We begin by creating our change predicates[6]. There are three change functions that are necessary for complete change predicates, we define them as follows:

- $Agreement(c, m) = (area(c \cap m))/(area(c \cup m))$
- $Containment(c, m) = (area(c \cap m))/(area(m))$
- $Overlap(s, p) = area(s \cap (p_1 \cap \ldots \cap p_m))/area(s)$

Here, c is a spatial cluster polygon from the batch timed t_i and m is another spatial cluster polygon from a batch timed t_{i+x}, with x indicating a time increase. We denote cluster intersection as $\cap$ and union as $\cup$. $Area(x)$ denotes the area covered by the polygon x, upto a desired level of accuracy.

The change function $Agreement(c, m)$ is a measure of similarity. $Containment(c, m)$ measures whether individual polygons coincide. To measure concurrence between one polygon and a set of polygons, we use the $Overlap(s, p)$ functions. Here $p_1, ..., p_m$ are spatial clusters from a batch b, and s is a cluster from a different batch b'. Other basic functions include $Centroid(x)$, which indicates the Centorid of a polygon as defined as:

$C_x = \frac{1}{6A} \sum_{i=0}^{N-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$
$C_y = \frac{1}{6A} \sum_{i=0}^{N-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$

where $x_N = x_0$ and A = the area of the polygon. We also use a simple Manhattan distance function $Distance(c, m)$ which is tuned to the required georeferenced accuracy. We use all these functions to defined these seven change predicates:

1) $S - Continuing(c, m) \leftrightarrow Agreement(c, m) \geq 0.8$
2) $B - Continuing(c, b) \leftrightarrow Overlap(c, b) \geq 0.8$
3) $Growing(c, m) \leftrightarrow Containment(c, m) \geq 0.9$
4) $Shrinking(c, m) \leftrightarrow Growing(m, c)$
5) $Disappearing(c) \leftrightarrow \exists i(belong - to(c, i)$
6) $Novel(c) \leftrightarrow \exists i(belong - to(c, i)and(i = 1ornot(B - Continuing(c, i - 1))$
7) $Shifting \leftrightarrow \frac{\sum_{i=1}^{N} Distance(Centroid(c), Centroid(m))}{N}$
   Implementations of these change predicates can be found in the Section IV.

B. Polygon Generation from Point Data Sources

We improve upon the change analysis framework called Aconcagua [6]. The system expects an input of emotion polygons annotated with emotion assessment scores, with +1 representing a very high positive emotion and -1 representing a very high negative emotion. While this method does lead to inconsistencies in locations due to georeferencing inaccuracies, we find that the 8   10m precision [8] works well within city limits.

There are several ways of using the numerous point data we have obtained from this step into an actual polygonal map:

1) Creating closed contour lines for contour lines that lie on the boundary of the observation area.
2) Creating a convex hull from points with similar scores.

We elaborate on creating convex hulls in the following section.

C. Change Analysis for Polygonal Map Data Sources

Our approach uses three primary set operations: union, intersection and erase. We calculate the area of each individual polygon within each map layer. We then execute a union operation and calculate area. The union layer now contains the original areas of both layers and the areas of the overlapping polygons - we now need to query them properly to prepare for calculating the change percentage and tabulating intersection.

To outline the polygon, we examine several different methods:

1) We find features common to either of the layers but not both, essentially performing a symmetrical difference
2) We erase the larger of the polygons from the smaller, thus retaining only the growth, and do vice-verse for shrinking
3) We perform simple intersection and then invert selection to get changed regions.

Our approach then combined several techniques:

1) **Data Pre Processing** This involves curation of datasets with obvious georeferencing errors. This would preferable be done by minimizing the root mean square error. We initially.
2) **Parametrization of polygons** Calculate shape and area parameters for each individual polygon with each map layer.s
3) **Analysis through Symmetrical Difference** Extract features common to either of the layers

4) **Polygon Union Computation** Union sequential layers to contain the original areas of both layers and the areas of overlapping polygons
5) **Polygon Erase Operation** Erase larger polygons from smaller (or vice-versa) for detection of growth/shrinking
6) **Polygon Intersection/Invert** Selecting and then labelling the changed regions

D. Storytelling Techniques for Geospatial Data

In order to tell a coherent spatio-temporal data story from the change analysis output, we propose using an interestingness function for this task.[11]

We choose a set of change polygons SCP. It not only contains those polygons but also their associated characteristics. For example, a SCP could contain a set of spatial clusters represented by polygons, their average drought score, total area of each polygon, centroid coordinates of each polygon and other summaries for each spatial cluster (polyogn). We define the function as:

$$f : SCP \ \rightarrow \ [0 \ , \infty)$$

We define a threshold ω, which ensures that a narrative will only be generated an object $p \in SCP$ such that $f(p) \geq \omega$. Sample parameters for ω include

- $Max\left(\frac{area(Polygon\ P_i)}{\sum_1^i area(Polygon\ P_i)}\right)$
- $Max\,(Percentage\ Change\ in\ Polygon)$
- Largest shift in polygon centroids

The threshold parameters need to be finely tuned so as to not exclude those polygons who fall through exceptions. Once we have a suitable selection of polygons and chose a threshold value, we can create a summary narrative based on that.

REFERENCES

[1] J. Im and J. Jensen. 2005. A change detection model based on neighborhood correlation image analysis and decision tree classification. Remote Sensing of Environment 99, 3 (2005), 326–340. DOI:http://dx.doi.org/10.1016/j.rse.2005.09.008

[2] Suming Jin, Limin Yang, Patrick Danielson, Collin Homer, Joyce Fry, and George Xian. 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. Remote Sensing of Environment 132 (May 2013), 159–175. DOI:http://dx.doi.org/10.1016/j.rse.2013.01.012

[3] D. Lu, P. Mausel, E. Brondízio, and E. Moran. 2004. Change detection techniques. International Journal of Remote Sensing 25, 12 (June 2004), 2365–2401. DOI:http://dx.doi.org/10.1080/0143116031000139863

[4] Merrill K. Ridd and Jiajun Liu. 1998. A Comparison of Four Algorithms for Change Detection in an Urban Environment. Remote Sensing of Environment 63, 2 (1998), 95–100. DOI:http://dx.doi.org/10.1016/s0034-4257(97)00112-0

[5] S. Minu and Amba Shetty. 2015. A Comparative Study of Image Change Detection Algorithms in MATLAB. Aquatic Procedia 4 (March 2015), 1366–1373. DOI:http://dx.doi.org/10.1016/j.aqpro.2015.02.177

[6] K. Elgarroussi, S. Wang, R. Banerjee, and C. F. Eick, “Aconcagua: A Novel Spatiotemporal Emotion Change Analysis Framework,” in Proceedings of the 2Nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, New York, NY, USA, 2018, pp. 54–61.

[7] R. Banerjee, K. Elgarroussi, S. Wang, A. Talari, Y. Zhang, and C. F. Eick, “K2: A Novel Data Analysis Framework to Understand US Emotions in Space and Time,” Int. J. Semantic Computing, vol. 13, no. 01, pp. 111–133, Mar. 2019.

[8] Geomenke, “How Accurate is the GPS on my Smart Phone? (Part 2),” Community Health Maps, 07-Jul-2014. .

[9] W. Z.-M.-L. I. for the S. S. Person, “Geotagged Twitter posts from the United States: A tweet collection to investigate representativeness,” Jan. 2016.

[10] C. J. Hutto, VADER Sentiment Analysis. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social .. 2019.

[11] X. Huang, A. An, and N. Cercone, “Comparison of Interestingness Functions for Learning Web Usage Patterns,” in Proceedings of the Eleventh International Conference on Information and Knowledge Management, New York, NY, USA, 2002, pp. 617–620.

[12] W. Wang, Lu Chen, K. Thirunarayan, and Amit P. Sheth, Harnessing Twitter ”Big Data” for Automatic Emotion Identifcation, in Proc. of the IEEE 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing.

[13] R. F. Dos Santos, S. Shah, F. Chen, A. Boedi-hardjo, P. Butler, C.T. Lu and N. Ramakrishnan, ”Spatiotemporal storytelling on twitter”, Virginia Tech Computer Science Technical Report, 2015, [online] Available: http://vtechworks.lib.vt.edu/handle/10919/24701. [20] D. Kumar, N. Ramakrishnan, R. F. Helm and M. Potts, Algorithm for storytelling, in IEEE Transactions on Knowledge and Data Engineering, 2008, Volume 20 Issue 6, pp. 736-751.

[14] N. Ramakrishnan, D. Kumar, B. Mishra, M. Potts and R. R. Helm, Turning CARTwheels: an alternating algorithm for mining redescriptions, in Knowledge Discovery and Data Mining, 2004, pp. 266-275.

[15] L. D. Vocht, C. Beecks, R. Verborgh, T. Seidl, E. Mannens and Rik Van De Walle, Improving Semantic Relatedness in Paths for Storytelling with Linked Data on the Web, in Revised Selected Papers of the ESWC 2015 Satellite Events on The Semantic Web: ESWC 2015 Satellite Events, May 31-June 04, 2015.

[16] Z. Chen, X. Zhang, A. P. Boedihardjo, J. Dai, Chang-Tien Lu, Multimodal storytelling via generative adversarial imitation learning, in Proceedings of the 26th International Joint Conference on Artifcial Intelligence, August 19-25, 2017, Melbourne, Australia.

[17] R. D. Santos, S. Shah, F. Chen, A. Boedihardjo, Chang-Tien Lu, N. Ramakrishnan, Forecasting location-based events with spatio-temporal storytelling, in Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, pp. 13-22, November 04-04, 2014, Dallas/Fort Worth, Texas.

[18] Jayant Madhavan, Sreeram Balakrishnan, Kathryn Brisbin, Hector Gonzalez, Nitin Gupta,Alon Halevy, Karen Jacqmin-Adams, Heidi Lam,

Anno Langen, Hongrae Lee,Rod McChesney, Rebecca Shapley, Warren Shen, “Big Data Storytelling through Interactive Maps”
[19] M. Shahriar Hossain1, Patrick Butler1, Arnold P. Boedihardjo2, Naren Ramakrishnan, Storytelling in Entity Networks to Support Intelligence Analysts, KDD ’12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, Pages 1375-1383 ,Beijing, China , August 12 - 16, 2012
[20] Zhiqian Chen , Xuchao Zhang , Arnold P. Boedihardjo , Jing Dai , Chang-Tien Lu, Multimodal storytelling via generative adversarial imitation learning, Proceedings of the 26th International Joint Conference on Artificial Intelligence, August 19-25, 2017, Melbourne, Australia
[21] Raimundo F. Dos Santos, Jr. , Arnold Boedihardjo , Sumit Shah , Feng Chen , Chang-Tien Lu , Naren Ramakrishnan, The big data of violent events: algorithms for association analysis using spatio-temporal storytelling, Geoinformatica, v.20 n.4, p.879-921, October 2016
[22] Deept Kumar, Naren Ramakrishnan, Richard F. Helm, and Malcolm Potts. 2008. Algorithms for Storytelling. IEEE TKDE20, 6 (2 2008), 736–751.DOI:http://dx.doi.org/10.1145/1188913.1188915
[23] “GRACE Tellus Data,” GRACE Tellus. [Online]. Available: https://grace.jpl.nasa.gov/data/get-data. [Accessed: 26-Jul-2019].
[24] “SARAL - eoPortal Directory - Satellite Missions.” [Online]. Available: https://directory.eoportal.org/web/eoportal/satellite-missions/s/saral. [Accessed: 26-Jul-2019].
[25] “Top 5 sources of big data — Artificial Intelligence — Data Science —,” 26-Nov-2017. [Online]. Available: https://www.allerin.com/blog/top-5-sources-of-big-data. [Accessed: 26-Jul-2019].
[26] J. Thatcher, “Big Data, Big Questions— Living on Fumes: Digital Footprints, Data Fumes, and the Limitations of Spatial Big Data,” International Journal of Communication, vol. 8, no. 0, p. 19, Jun. 2014.
[27] “Livehoods.” [Online]. Available: http://livehoods.org/research. [Accessed: 27-Jul-2019].
[28] “Shapefiles of drought intensity and impact for the North American Drought Portal.” [Online]. Available: https://www1.ncdc.noaa.gov/pub/data/nidis/shapefiles/. [Accessed: 01-Aug-2019].