

# Approach to Automated Storytelling Using Geospatial-Temporal 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

Our change predicates are derived from the Aconcagua Framework [6]. We consider  $c$  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.

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

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  $p_1, \dots, p_m$  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)  $Novel(c) \leftrightarrow \exists i(belong - to(c, i) \text{ and } (i = 1 \text{ or not}(B - Continuing(c, i - 1)))$
- 7)  $Shifting \leftrightarrow \frac{\sum_{i=1}^N Distance(Centroid(c), Centroid(m))}{N}$

### B. Interestingness Functions

In order to tell a coherent spatio-temporal data story from the change analysis output, we propose using an interestingness function for this task.[1]

We choose an area of interest  $A(x_1, y_1, x_2, y_2)$  which is a rectangle bounded by coordinates  $(x_1, y_1)$  and  $(x_2, y_2)$ . The length of the diagonal is  $d$ . We choose a set of polygons SCP, whose centroid  $C_{x,y}$  lies within  $A$ .

The Thinness ratio measure takes a maximum value of 1 for a circle. Objects of regular shape have a higher thinness ratio than similar irregular ones. We define the thinness ratio of a polygon as:

$$T = \frac{4\pi Area(SCP)}{Perimeter^2}$$

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 (polygon). We define the function as:

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

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

- $\alpha = (Max(\frac{area(Polygon P_i)}{\sum_1^i area(Polygon P_i)}))$
- $\beta = (Max(Percentage Change in Polygon))$
- $\gamma = (Shifting \geq \frac{d}{2})$
- $\delta = (T \geq 0.3)$

The parameters  $\beta$  and  $\gamma$  are special cases.  $\beta$  is to be used for those SCP whose primary change was growth/shrinking, henceforth referred to as  $SCP_{ext}$ .  $\gamma$  is to be used for those SCP whose primary change was shifting, henceforth referred to as  $SCP_{shift}$ . 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.

Thus, our Interestingness function  $\omega$  is hence calculated as:

$$\omega_{ext} = 0.3\alpha + 0.4\beta + 0.4\delta$$

$$\omega_{shift} = 0.3\alpha + 0.4\gamma + 0.4\delta$$

## REFERENCES

- [1] 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.