Visual Analytics of Time Series Data

Chris Zhenghao Chen | Summer Scholar 2016
February 2016

www.data61.csiro.au

Stage 1 Preferences Clusters

- Objectives
  - Discover preferences of certain attractions[1]
  - Group customers with similar preferences[1]
  - Visualize trajectory patterns[2]

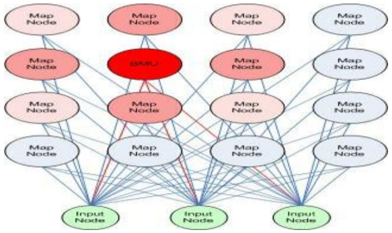
- Data Preprocessing
  - Check-in Frequency Data
  - Time-spend Data
  - Attraction Data

Stage 1 Preferences Clusters

Methodologies

- K-Nearest Neighborhoods[3] & K-means[1][4]
  - Cluster top-k groups of customers[3][4]
  - Supervisory, performance depends on k, sensitive to noise [4]

- Self-Organizing Map[5][11]
  - Un-supervisory, Artificial Neuron Network [5], extension of K-means [6]
  - Architecture[5]



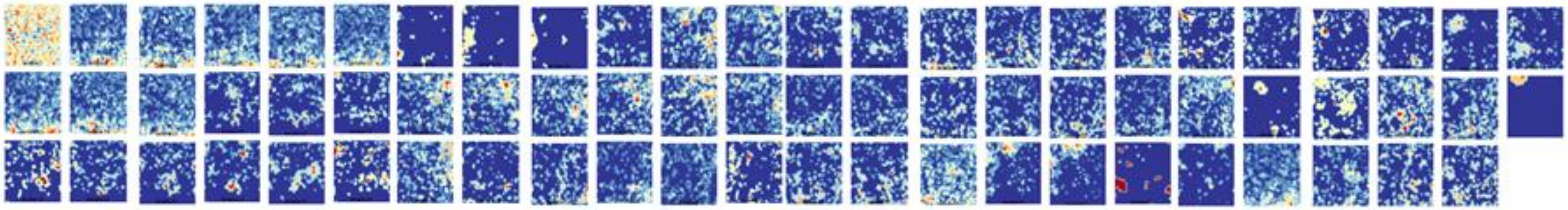
  - Algorithm[5]

Input → Training
- Initialization
- Iteration
  - Competition $Dist = \sqrt{\sum_{i=0}^{i=n} (V_i - W_i)^2}$
  - Cooperation $\sigma(t) = \sigma_0 e^{\left(-\frac{t}{\lambda}\right)}$ t=1, 2, 3 ....
  - Adaptation $W(t+1) = W(t) + L(t)(V(t)-W(t))$
    - $L(t)=L_0 e^{\left(-\frac{t}{\lambda}\right)}$ t= 1, 2, 3.....
    - $\Theta(t) = e^{\left(-\frac{dist^2}{2\sigma^2(t)}\right)}$ t=1, 2, 3...

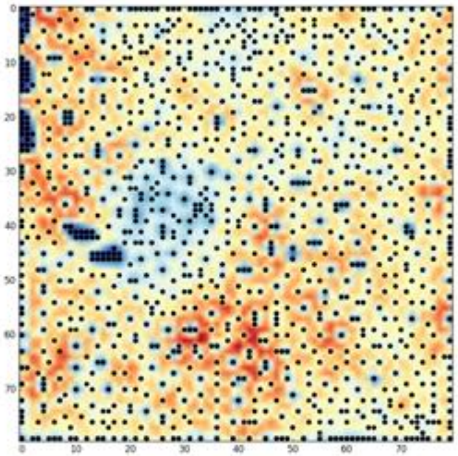
→Render

Self-organizing Map Results in Stage 1

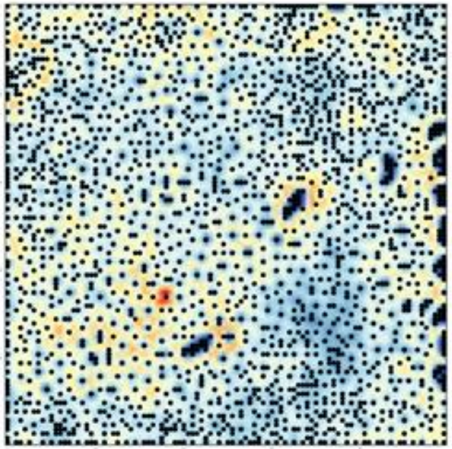
- Heat Map for single Attraction



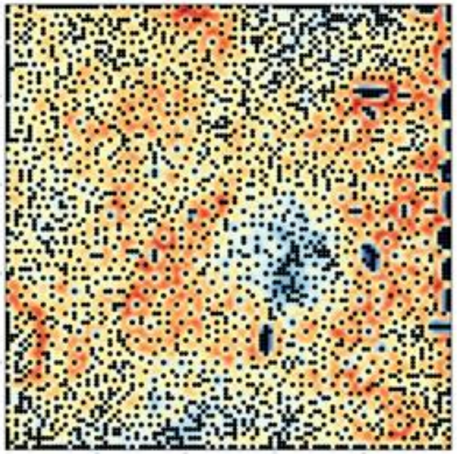
- U-matrix for clustering preference of attraction 3 days



Friday
Map size 80*80
1000 times training



Saturday
Map Size: 100*100
1000 times training

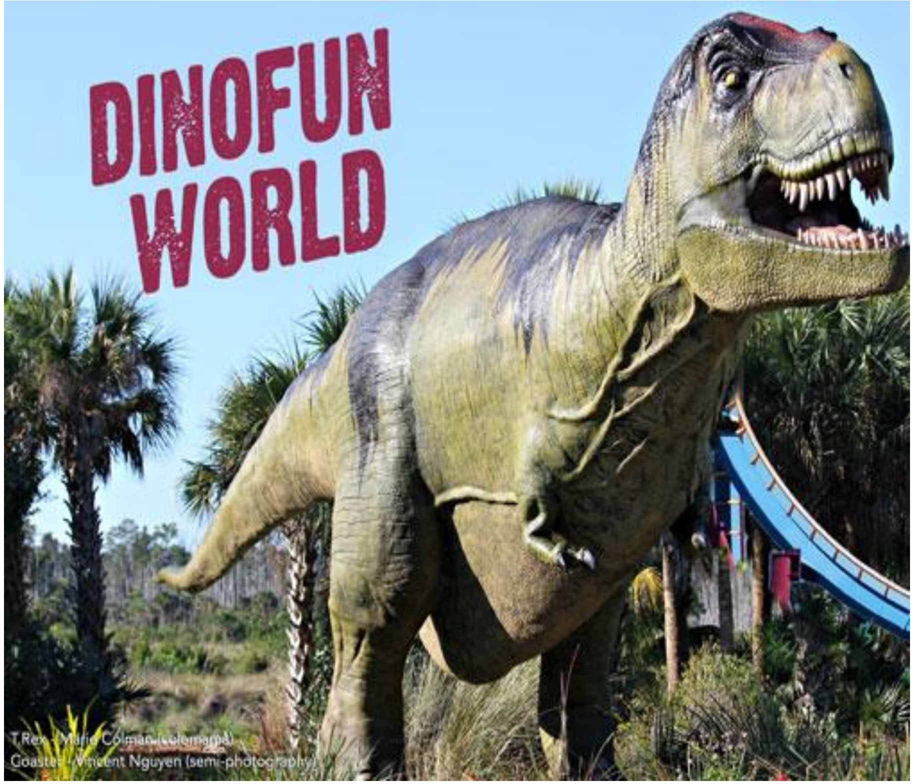


Sunday
Map Size: 100*100
1000 times training

Stage 2 Groups Visualization

- Objectives
  - Discover trajectory of groups
  - Find difference[6] & similarity[8] of groups
  - Visualize the overall similarity[3]
  - Study communications pattern[1][2]

- Data Preprocessing
  - Groups Data (From Stage 1)
  - Aggregated Trajectory Data
  - Trajectory Data for Single Person
  - Average Speed Rates Change Data
  - Communication Data



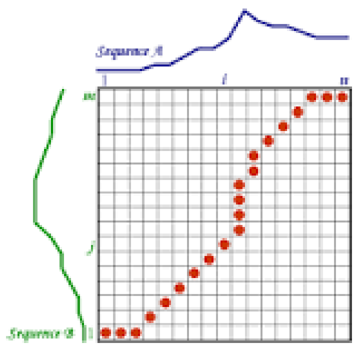
Stage 2 Groups Visualization

Methodologies

- Frequency & Time Spend Visualization
  - Bar chart & Heat map[1]
- Real Time Trajectory Computation
  - Using Animation[7] & 3D plot
- Communication Pattern Visualization[1][2]
  - Line Chart
- Dynamic Time Warping[9]
  - Calculate similarity of 2 Sequence Data[10]
  - Algorithm[10]

$$g(\emptyset, \emptyset) = 0$$
$$g(s, \emptyset) = dist(\emptyset, q) = \infty$$
$$g(s, q) = dist(s_1, q_1) + min \left\{ \begin{array}{l} g(s, \langle q_2, \ldots, q_m \rangle) \\ g(\langle s_2, \ldots, s_m \rangle, q) \\ g(\langle s_2, \ldots, s_m \rangle, \langle q_2, \ldots, q_m \rangle) \end{array} \right\}$$

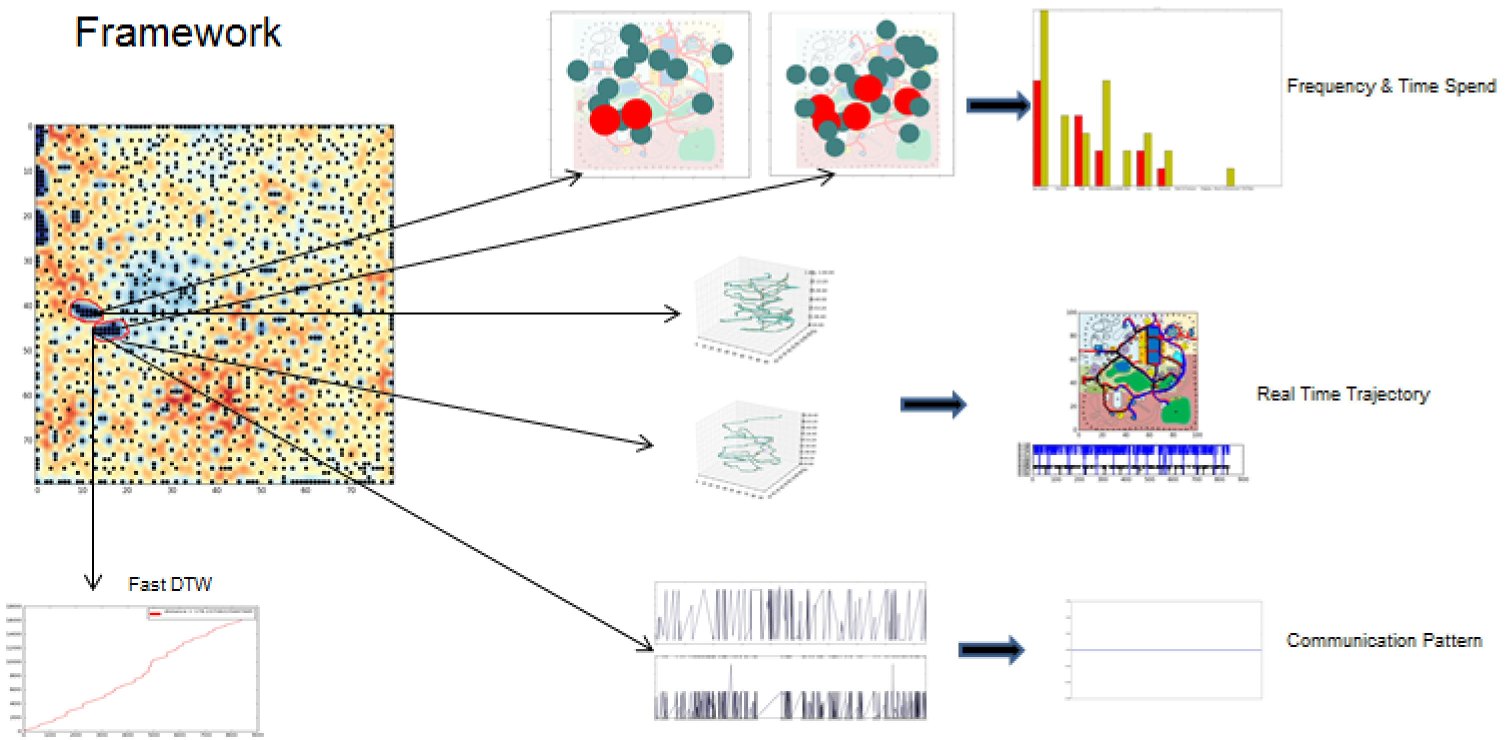
  - Running Time [9] & Memory Space[9] $O(n^2)$



- Fast Dynamic Time Warping[10]
  - A Optimization of Dynamic Time Warping
  - Running time[10] & Memory Space[10] $O(n)$
- Distance Matrix[3]
  - Distance Value ***D = D1*W1 + D2*W2 + …. Dn*Wn***

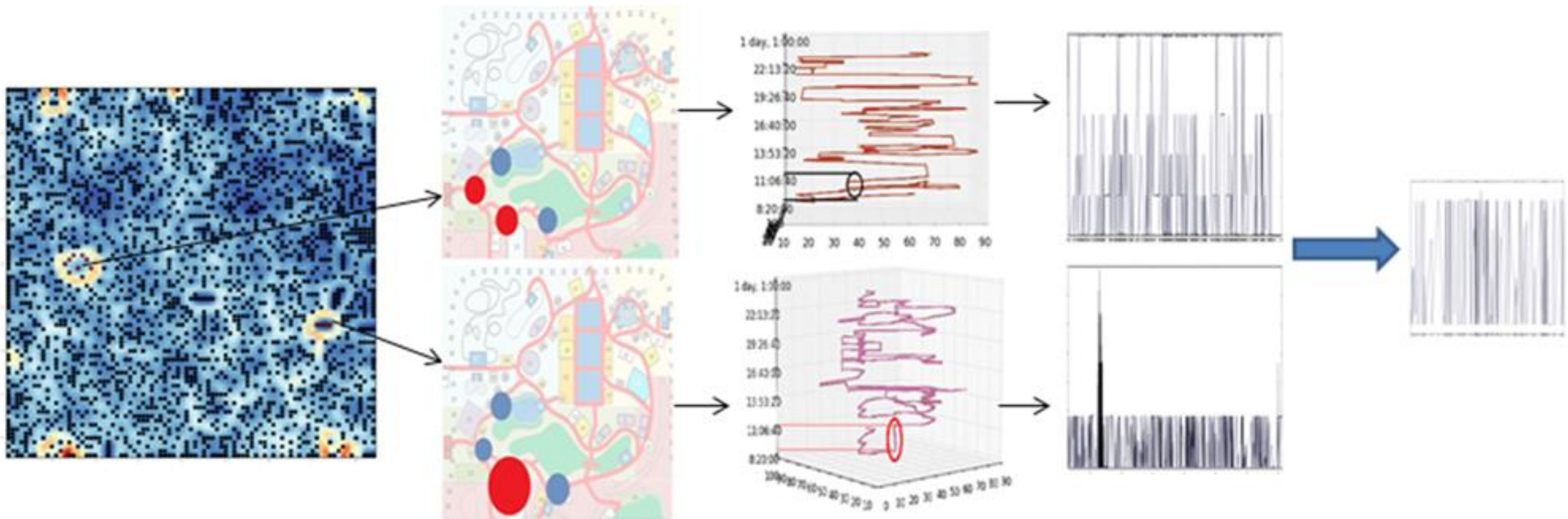
THE UNIVERSITY OF SYDNEY

Framework of Stage 2



Case Study: Crime Detection

- Background
  - A Crime Group vandalized exhibiting of a local soccer star(DinoFunNews 2014)
  - Occurred between 9:40 AM to 11:30 AM[2]
- Crime Visualization

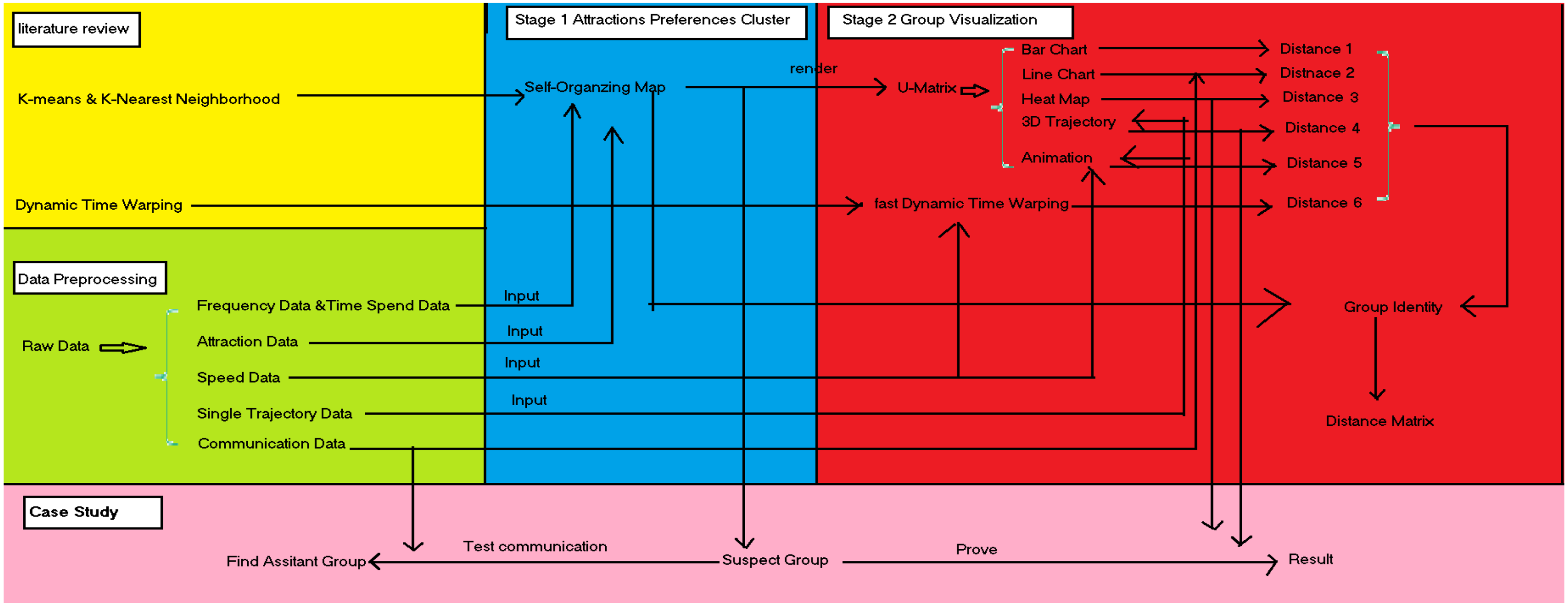


- Conclusion

Group 2 has high probability to commit while further hypothesize Group 1 assist crime

Project Summary and Further Work

- Project Summary



- Future Work
  - Trajectory can be further clustered (e.g. hierarchical Dirichlet process Hidden Markov Model)
  - Behaviour and motivation can be learnt and prediction (e.g. Deep Learning)

Reference

[1] ParkAnalyzer: Characterizing the Movement Patterns of Visitors. VAST 2015 Mini-Challenge 1. Jieqiong Zhao. Guizhen Wang. Junghoon Chae. Hanye Xu.

[2] Steptoe, Michael, et al. "VAST Challenge 2015: Grand Challenge-Team VADER/VIS Award for Outstanding Comprehensive Submission." *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*. IEEE, 2015.

[3] Zheng, Yu. "Trajectory data mining: an overview." *ACM Transactions on Intelligent Systems and Technology (TIST)* 6.3 (2015): 29.

[4] Ossama, Omnia, Hoda MO Mokhtar, and Mohamed E. El-Sharkawi. "An extended k-means technique for clustering moving objects." *Egyptian Informatics Journal* 12.1 (2011): 45-51.

[5] Kohonen, T. (1998). The self-organizing map. Neurocomputing, 21(1-3), pp.1-6.

[6] Baçao, Fernando, Victor Lobo, and Marco Painho. "Self-organizing maps as substitutes for k-means clustering." *Computational Science–ICCS 2005*. Springer Berlin Heidelberg, 2005. 476-483.

[7] Visual Analysis of Route Choice Behaviour based on GPS Trajectories. Min Lu1,2,3. Chufan Lai1,2,3. Tangzhi Ye1,2. Jie Liang1,2. Xiaoru Yuan1,2,3 ∗.

[8] Haase, Jens, and Ulf Brefeld. "Finding similar movements in positional data streams." *Proc. ECML/PKDD Workshop on Machine Learning and Data Mining for Sports Analytics*. 2013.

[9] Müller, Meinard. "Dynamic time warping." *Information retrieval for music and motion* (2007): 69-84.

[10] Salvador, Stan, and Philip Chan. "Toward accurate dynamic time warping in linear time and space." *Intelligent Data Analysis* 11.5 (2007): 561-580.

[11] Moosavi, Vahid. "Computing With Contextual Numbers." *arXiv preprint arXiv:1408.0889* (2014).

Thank you

Data61 | Machine Learning Group
Chris Zhenghao Chen
Summer Scholar 2016

t +61 04 1657 1168
E Zhenghao.Chen@nicta.com.au