CIKM 2022

Mining Reaction and Diffusion Dynamics in Social Activities

Taichi Murayama, Yasuko Matsubara, Yasushi Sakurai

SANKEN, Osaka University

Online User Activity on the Web as Sensor







Physics World

Social sensor

Example: Google Flu

A system that predicts the number of influenza cases based on the volume of search queries based on linear regression



Sources: http://www.google.org/fluttends/us. CDC ILInet data from http://gis.cdc.gov/grasp/fluview/fluportaldas/board.html, Cook et al. (2011) Assessing Google Flu Trends Performance in the United States during the 2009 influenza Virus A (H1N1) Pandemic. PLoS ONE 6(8): e23810. doi:10.1371/journal.pone.0023810. Data as of Jan. 12, 2013. Keith Winstein (keithw@mit.edu)

Search queries capture human interests



Given: online user activities



Goal 1: Diffusion flow of each group



Goal 2: Interaction between keywords



Goal 3: Seasonality





Goal 4: Good forecasting performance



Given: online user activities

Goal 1: Diffusion process of each group Goal 2: Interaction between keywords Goal 3: Seasonality Goal 4: Good forecasting performance

Given: online user activities

Goal 1: Diffusion process of each group Goal 2: Interaction between keywords Goal 3: Seasonality Goal 4: Good forecasting performance

Our solution: FLUXCUBE Reaction-diffusion system + Neural Network

Application of Reaction-Diffusion System Main Equation $\frac{\partial u}{\partial t} = \mathbf{f}(u,t) + \mathcal{D}\Delta u$ Diffusion term **Reaction term Reaction-Diffusion System is utilized is** a mathematical model corresponding for physical phenomena.







Input: $\boldsymbol{\chi} = \{\boldsymbol{x}_{tij}\}$

timestep t, location i, item j

FLUXCUBE



$$x_{t+1ij} = F\left(\frac{\partial x_{tij}}{\partial t} + x_{tij}\right)$$





Component idea: Item interaction as Jungle



The idea represents as Lotka-Volterra Equation

$$a_j \boldsymbol{\chi}_{t,i,j} \left(1 - \frac{\sum_{j} c_{jj} \boldsymbol{\chi}_{tij}}{b_j} \right)$$

- *a_j*: intrinsic growth rate of keyword *j*;
- *b_j*: carrying capacity of keyword *j*;
- $c_{jj'}$: intra/inter-keyword interaction strength from the j'-th keyword to the j-th keyword, which is each value in C



The idea represents as Lotka-Volterra Equation

$$a_j \boldsymbol{\chi}_{t,i,j} \left(1 - \frac{\sum_{j} c_{jj} \boldsymbol{\chi}_{tij}}{b_j} \right)$$

The variable c_{jj} , represents the kinds of interaction between item j and item j'

Adults

- $c_{jj'} > 0, c_{j'j} > 0$: a competitive relationship
- $c_{jj'} < 0, c_{j'j} > 0$: a parasitic relationship
- $c_{jj'} < 0, c_{j'j} = 0$: a commensal relationship
- $c_{jj'} < 0, c_{j'j} < 0$: a mutualistic relationship

Kids

Teens

The idea represents as Lotka-Volterra E

$$\alpha_j \boldsymbol{\chi}_{t,i,j} \left(1 - \frac{\sum_{j} c_{jj} \boldsymbol{\chi}_{j}}{b_j} \right)$$

• $c_{jj'} > 0, c_{j'j} > 0$: a competitive relationship

• $c_{jj'} < 0, c_{j'j} = 0$: a commensal relationship

• $c_{jj'} < 0, c_{j'j} < 0$: a mutualistic relationship

• $c_{jj'} < 0, c_{j'j} > 0$: a parasitic relationship

$$b_j \begin{pmatrix} c_{jj}, x_{tij}, \\ b_j \end{pmatrix}$$

Spider monkeys

Fruits

The idea represents as Lotka-Volt

$$a_j \boldsymbol{\chi}_{t,i,j} \left(1 - \frac{\sum_{j'} c_{jj'} \boldsymbol{\chi}_{tij'}}{b_j} \right)$$
 be

<text>

- $c_{jj'} > 0, c_{j'j} > 0$: a competitive relationship
- $c_{jj'} < 0, c_{j'j} > 0$: a parasitic relationship
- $c_{jj'} < 0, c_{j'j} = 0$: a commensal relationship
- $c_{jj'} < 0, c_{j'j} < 0$: a mutualistic relationship



The idea represents as Lotka-Volterra Equation

$$a_j \chi_{t,i,j} \left(1 - \frac{\sum_{j'} c_{jj'} \chi_{tij'}}{b_j} \right)^{\text{Theorem 1}}$$

- $c_{jj'} > 0, c_{j'j} > 0$: a competitive relationship
- $c_{jj'} < 0, c_{j'j} > 0$: a parasitic relationship
- $c_{jj'} < 0, c_{j'j} = 0$: a commensal relationship
- $c_{jj'} < 0, c_{j'j} < 0$: a mutualistic relationship



The idea represents as Lotka-Volterra Equation

FLUXCUBE:
$$\frac{\partial x_{tij}}{\partial t} = \int (x_{tij} | \chi_{t,i,:}) + \mathcal{G}(x_{tij} | \chi_t, t)$$

$$\int f(x_{tij} | \chi_{t,i,:}) = a_j \chi_{t,i,j} \left(1 - \frac{\sum_{j, i} c_{jj, i} \chi_{tij,i}}{b_j}\right)$$



Reaction-Diffusion
$$\frac{\partial u}{\partial t} = f(u, t) + D\Delta t$$

Diffusion term
 $\frac{\partial x_{tij}}{\partial t} = f(x_{tij}|\chi_{t,i,:}) + g(x_{tij}|\chi_t, t)$
Seasonal term

$$x_{t+1ij} = F\left(\frac{\partial x_{tij}}{\partial t} + x_{tij}\right)$$

- The interactions of any keyword between locations are not constant because of external factors
- We need to capture the time change of the interaction, e.g., complex phenomena and rapid changes.



- The interactions of any keyword between locations are not constant because of external factors
- We need to capture the time change of the interaction, e.g., complex phenomena and rapid changes.

Our model represents the interaction by RNN



Our model represents the interaction between locations by **Recurrent Neural Network**

By applying a neural network to apart of our mathematical model, we expect to achieve both **flexible modeling and high explainability.**

FLUXCUBE:
$$\frac{\partial x_{tij}}{\partial t} = f(x_{tij}|\boldsymbol{\chi}_{t,i,:}) + \mathcal{G}(x_{tij}|\boldsymbol{\chi}_{t}, t)$$
$$\mathcal{G}(x_{tij}|\boldsymbol{\chi}_{t}, t) = \sum RNN(1:t) \odot \boldsymbol{\chi}_{t}$$

The output value of RNN represents the contribution of the popularity of each keyword between locations.





Reaction-Diffusion
$$\frac{\partial u}{\partial t} = f(u, t) + D\Delta u$$

Seasonlaity term
 $\frac{\partial x_{tij}}{\partial t} = f(x_{tij}|\chi_{t,i,:}) + g(x_{tij}|\chi_t, t)$

$$x_{t+1ij} = F\left(\frac{\partial x_{tij}}{\partial t} + x_{tij}\right)$$

Finding hidden seasonality

Online users change their behavior according to seasonal events, such as Christmas and Black Friday.







FLUXCUBE



$$x_{t+1ij} = F\left(\frac{\partial x_{tij}}{\partial t} + x_{tij}\right)$$

Regression termSparse termLoss Function
$$\||\chi^c - \hat{\chi}||^2 + \alpha \sum ||D||^2 + \beta \sum ||S||^2$$

 It is difficult to infer interactions between many areas due to computational costs. It is difficult to infer interactions between many areas due to computational costs.

Our solution: Before training, grouping similar areas into same group



Reaction term

FLUXCUBE:
$$\frac{\partial x_{tij}}{\partial t} = \int (x_{tij} | \boldsymbol{\chi}_{t,i,:}) + g(x_{tij} | \boldsymbol{\chi}_{t}, t)$$
$$f(x_{tij} | \boldsymbol{\chi}_{t,i,:}) = a_j \boldsymbol{\chi}_{t,i,j} \left(1 - \frac{\sum_{j} c_{jj} x_{tij}}{b_j}\right)$$
$$\begin{pmatrix} g^1, \dots, g^d \end{pmatrix} = \text{K-means} \left(\text{UMAP}(a^i, b^i, C^i)\right), \quad i = [1, \dots, L]$$
Grouping each areas Parameters in Reaction term of each country Number of countries

Country Grouping



$$g^1, \dots g^{d^l}$$
=K-means $\left(\text{UMAP}(a^i, b^i, C^i) \right), i = [1, \dots, L]$ Grouping
each areasParameters in Reaction term
of each countryNumber of
countries

Experimental Settings

- Evaluate the forecasting score in 13, 26, 52 weeks ahead
- Evaluation Metrics
 - RMSE, MAE: Smaller value indicating better performance

Dataset

 Two types of Google Trend data, which contained weekly web search volumes collected for about 10 years

ID	Dataset	Query						
US#1	E-commerce	Amazon/Apple/BestBuy/Costco/Craigslist/Ebay/ Homedepot/Kohls/Macys/Target/Walmart						
US#2	VoD	AppleTV/ESPN/HBO/Hulu/Netflix/Sling/ Vudu/YouTube						
US#3	Sweets	Cake/Candy/Chocolate/Cookie/Cupcake/ Gum/Icecream/Pie/Pudding						
US#4	Facilities	Aquarium/Bookstore/Gym/Library/Museum/ Theater/Zoo						
World#1	Music	Beyonce/KatyPerry/LadyGaga/Maroon5/ StevieWonder/TaylorSwift						
World#2	SNS	Facebook/LINE/Slack/Snapchat/Twitter/ Viber/WhatsApp						
World#3	Apparel	Gap/H&M/Primark/Uniqlo/Zara						

Experimental Performance

		13 w	reeks	26 weeks 52 week		eeks			13 weeks		26 weeks		52 weeks		
Dataset	Model	RMSE	MAE	RMSE	MAE	RMSE	MAE	Dataset	Model	RMSE	MAE	RMSE	MAE	RMSE	MAE
	EcoWeb	0.1470	0.0950	0.1554	0.1082	0.1654	0.1197		EcoWeb	0.1259	0.0831	0.1422	0.1034	0.2101	0.1460
	SMF	0.0869	0.0620	0.0910	0.0654	0.1012	0.0674		SMF	0.0936	0.0783	0.0901	0.0602	0.1087	0.0787
	DeepAR	0.1003	0.0634	0.1302	0.0907	0.1385	0.1014	World#1	DeepAR	0.0900	0.0636	0.0929	0.0681	0.1395	0.0972
US#1	GRU	0.1723	0.1175	0.1924	0.1374	0.2059	0.1525		GRU	0.0633	0.0452	0.0718	0.0501	0.0823	0.0572
	Informer	0.1477	0.1045	0.1375	0.0985	0.1575	0.1111		Informer	0.0704	0.0423	0.0719	0.0416	0.0738	0.0446
	FluxCube	0.0478	0.0257	0.0574	0.0323	0.0631	0.0365		FluxCube	0.0454	0.0274	0.0477	0.0286	0.0546	0.0331
	EcoWeb	0.1440	0.1133	0.1981	0.1621	0.1920	0.1684	4 9 4 World#2 2	EcoWeb	0.0908	0.0300	0.1089	0.0570	0.1353	0.0742
	SMF	0.0621	0.0445	0.0713	0.0522	0.0760	0.0529		SMF	0.0841	0.0436	0.0799	0.0454	0.0826	0.0480
	DeepAR	0.1471	0.1026	0.1781	0.1314	0.1906	0.1474		DeepAR	0.0374	0.0098	0.0585	0.0199	0.0643	0.0209
US#2	GRU	0.1518	0.1171	0.1619	0.1231	0.1683	0.1440		GRU	0.0401	0.0159	0.0588	0.0174	0.0739	0.0254
	Informer	0.1277	0.0878	0.1292	0.0876	0.1436	0.1012		Informer	0.0371	0.0159	0.0595	0.0196	0.0642	0.0208
	FluxCube	0.0245	0.0130	0.0276	0.0156	0.0310	0.0181		FluxCube	0.0704	0.0271	0.0711	0.0304	0.0831	0.0351
	EcoWeb	0.1555	0.1208	0.1730	0.1384	0.1754	0.1369		EcoWeb	0.0523	0.0208	0.0626	0.0200	0.1080	0.0293
	SMF	0.0276	0.0186	0.0281	0.0170	0.0281	0.0190		SMF	0.0206	0.0111	0.0289	0.0160	0.0254	0.0195
110 110	DeepAR	0.1107	0.0753	0.1267	0.0833	0.1309	0.0908	World#3	DeepAR	0.0211	0.0110	0.0275	0.0099	0.0613	0.0214
US#3	GRU	0.1300	0.0869	0.1368	0.9843	0.1368	0.0939		GRU	0.0191	0.0090	0.0217	0.0096	0.00235	0.0115
	Informer	0.1322	0.0954	0.1311	0.0946	0.1279	0.0914		Informer	0.0223	0.0105	0.0226	0.0105	0.0214	0.0108
	FluxCube	0.0200	0.0121	0.0222	0.0136	0.0238	0.0148		FluxCube	0.0176	0.0085	0.0214	0.0096	0.0221	0.0100

Experimental Performance

		13 w	veeks	26 w	eeks	52 weeks				13 weeks		26 weeks		52 weeks	
Dataset	Model	RMSE	MAE	RMSE	MAE	RMSE	MAE	Dataset	Model	RMSE	MAE	RMSE	MAE	RMSE	MAE
US#1	EcoWeb	0.1470	0.0950	0.1554	0.1082	0.1654	0.1197		EcoWeb	0.1259	0.0831	0.1422	0.1034	0.2101	0.1460
	SMF	0.0869	0.0620	0.0910	0.0654	0.1012	0.0674		SMF	0.0936	0.0783	0.0901	0.0602	0.1087	0.0787
	DeepAR	0.1003	0.0634	0.1302	0.0907	0.1385	0.1014	World#1	DeepAR	0.0900	0.0636	0.0929	0.0681	0.1395	0.0972
	GRU	0.1723	0.1175	0.1924	0.1374	0.2059	0.1525		GRU	0.0633	0.0452	0.0718	0.0501	0.0823	0.0572
	Informer	0.1477	0.1045	0.1375	0.0985	0.1575	0.1111		Informer	0.0704	0.0423	0.0719	0.0416	0.0738	0.0446
	FluxCube	0.0478	0.0257	0.0574	0.0323	0.0631	0.0365		FluxCube	0.0454	0.0274	0.0477	0.0286	0.0546	0.0331
	EcoWeb	0.1440	0.1133	0.1981	0.1621	0.1920	0.1684	World#2	EcoWeb	0.0908	0.0300	0.1089	0.0570	0.1353	0.0742
	SMF	0.0621	0.0445	0.0713	0.0522	0.0760	0.0529		SMF	0.0841	0.0436	0.0799	0.0454	0.0826	0.0480
110	DeepAR	0.1471	0.1026	0.1781	0.1314	0.1906	0.1474		DeepAR	0.0374	0.0098	0.0585	0.0199	0.0643	0.0209
US#2	GRU	0.1518	0.1171	0.1619	0.1231	0.1683	0.1440		GRU	0.0401	0.0159	0.0588	0.0174	0.0739	0.0254
	Informer	0.1277	0.0878	0.1292	0.0876	0.1436	0.1012		Informer	0.0371	0.0159	0.0595	0.0196	0.0642	0.0208
	FluxCube	0.0245	0.0130	0.0276	0.0156	0.0310	0.0181		FluxCube	0.0704	0.0271	0.0711	0.0304	0.0831	0.0351
	EcoWeb	0.1555	0.1208	0.1730	0.1384	0.1754	0.1369		EcoWeb	0.0523	0.0208	0.0626	0.0200	0.1080	0.0293
	SMF	0.0276	0.0186	0.0281	0.0170	0.0281	0.0190		SMF	0.0206	0.0111	0.0289	0.0160	0.0254	0.0195
US#3	DeepAR	0.1107	0.0753	0.1267	0.0833	0.1309	0.0908	World#3	DeepAR	0.0211	0.0110	0.0275	0.0099	0.0613	0.0214
	GRU	0.1300	0.0869	0.1368	0.9843	0.1368	0.0939		GRU	0.0191	0.0090	0.0217	0.0096	0.00235	0.0115
	Informer	0.1322	0.0954	0.1311	0.0946	0.1279	0.0914		Informer	0.0223	0.0105	0.0226	0.0105	0.0214	0.0108
	FluxCube	0.0200	0.0121	0.0222	0.0136	0.0238	0.0148		FluxCube	0.0176	0.0085	0.0214	0.0096	0.0221	0.0100

Experimental Performance

		13 weeks		26 weeks 52 weeks		eeks			13 weeks		26 weeks		52 weeks		
Dataset	Model	RMSE	MAE	RMSE	MAE	RMSE	MAE	Dataset	Model	RMSE	MAE	RMSE	MAE	RMSE	MAE
US#1	EcoWeb	0.1470	0.0950	0.1554	0.1082	0.1654	0.1197		EcoWeb	0.1259	0.0831	0.1422	0.1034	0.2101	0.1460
	SMF	0.0869	0.0620	0.0910	0.0654	0.1012	0.0674		SMF	0.0936	0.0783	0.0901	0.0602	0.1087	0.0787
	DeepAR	0.1003	0.0634	0.1302	0.0907	0.1385	0.1014	World#1	DeepAR	0.0900	0.0636	0.0929	0.0681	0.1395	0.0972
	GRU	0.1723	0.1175	0.1924	0.1374	0.2059	0.1525		GRU	0.0633	0.0452	0.0718	0.0501	0.0823	0.0572
	Informer	0.1477	0.1045	0.1375	0.0985	0.1575	0.1111		Informer	0.0704	0.0423	0.0719	0.0416	0.0738	0.0446
	FluxCube	0.0478	0.0257	0.0574	0.0323	0.0631	0.0365		FluxCube	0.0454	0.0274	0.0477	0.0286	0.0546	0.0331
	EcoWeb	0.1440	0.1133	0.1981	0.1621	0.1920	0.1684	World#2	EcoWeb	0.0908	0.0300	0.1089	0.0570	0.1353	0.0742
	SMF	0.0621	0.0445	0.0713	0.0522	0.0760	0.0529		SMF	0.0841	0.0436	0.0799	0.0454	0.0826	0.0480
110	DeepAR	0.1471	0.1026	0.1781	0.1314	0.1906	0.1474		DeepAR	0.0374	0.0098	0.0585	0.0199	0.0643	0.0209
US#2	GRU	0.1518	0.1171	0.1619	0.1231	0.1683	0.1440		GRU	0.0401	0.0159	0.0588	0.0174	0.0739	0.0254
	Informer	0.1277	0.0878	0.1292	0.0876	0.1436	0.1012		Informer	0.0371	0.0159	0.0595	0.0196	0.0642	0.0208
	FluxCube	0.0245	0.0130	0.0276	0.0156	0.0310	0.0181		FluxCube	0.0704	0.0271	0.0711	0.0304	0.0831	0.0351
	EcoWeb	0.1555	0.1208	0.1730	0.1384	0.1754	0.1369		EcoWeb	0.0523	0.0208	0.0626	0.0200	0.1080	0.0293
	SMF	0.0276	0.0186	0.0281	0.0170	0.0281	0.0190		SMF	0.0206	0.0111	0.0289	0.0160	0.0254	0.0195
110	DeepAR	0.1107	0.0753	0.1267	0.0833	0.1309	0.0908	World#3	DeepAR	0.0211	0.0110	0.0275	0.0099	0.0613	0.0214
US#3	GRU	0.1300	0.0869	0.1368	0.9843	0.1368	0.0939		GRU	0.0191	0.0090	0.0217	0.0096	0.00235	0.0115
	Informer	0.1322	0.0954	0.1311	0.0946	0.1279	0.0914		Informer	0.0223	0.0105	0.0226	0.0105	0.0214	0.0108
	FluxCube	0.0200	0.0121	0.0222	0.0136	0.0238	0.0148		FluxCube	0.0176	0.0085	0.0214	0.0096	0.0221	0.0100



Experiments



Case study1: Vod



Keyword List

AppleTV / ESPN / HBO / Hulu / Netflix Sling / Vudu / Youtube





Case study1: Vod Influential flow







Case study2: Music

Keyword List

Beyonce/KatyPerry/LadyGaga/Maroon5/ StevieWonder/TaylorSwift





Timo



- FLUXCUBE: an effective modeling and forecasting method based on reaction-diffusion and ecological systems. It can recognize trends, seasonality and interactions in input observations by extracting their latent dynamic systems.
- Proposed model achieves higher accuracy in Google Trends Datasets by capturing the latent dynamics.
- It provides the latent interactions and the influence flows hidden behind observational data in a human-interpretable form.