A Multidimensional Method to Find Irregular Patterns in Interaction Data

Audrey Wilmet, Robin Lamarche-Perrin - first.lastname@lip6.fr
LIP6 - Sorbonne Université - Complex Network Team, France

Objectives and Approach

Goal: find unexpected behaviors related to political communication via Twitter.

In temporal interaction data, irregular patterns can be defined in an almost unlimited number of ways by differently combining time and structure at different levels.

An Irregular Pattern = an entity which statistically deviates from others [1]

1st Level: Choose an entity e
3D: (s, a, t)
2D: (s, a), (s, t), (a, t)
1D: a, s, t

ex: e = (s, a) ⇒ unusual relationships between spreaders and authors.
+communities (s, c, t)
+keywords (s, a, k, t)
+temporal granularity t = (d, h)

2nd Level: Measure its activity v

v(e) = number of retweets of e
ex: number of time s retweeted α
v(s, a, t) = ∑ v(s, a, t)

+Proportion: ex: v(s, a, t)/v(s) = proportion of retweets of s towards a among all retweets of s.

3rd Level: Choose the context

- Normalisation of the activity v by a more aggregated value [3].
  ex: observed vs v expected

- Selection of the set of entities which forms the normality.
  ex: e = (s, a) ⇒ Global: e ∈ V × V → Local: e ∈ V × c

Define What An Irregular Pattern Is

Perform Operations on the Data Cube

Case Study

1st step: Find Abnormal Hours

⇒ couples t = (d, h) during which the proportion of retweets observed during hour h of day d among all retweets of day d is higher than the expected proportion during hour h on all other days, compared to all other couples (d, h).

Entity = (d, h)
Context: Coverage
Activity: v_{obs} = v_{(a, d, h)} / v_{(a, d, h)}
Global: Days × Hours = [0, 31] × [0, 23]

ex: On the 24th of August, there has been much more retweets from 20:00 to 21:00 than usually are during this hour when looking at other days.

2nd step: Find Who Is Responsible For It

⇒ triplets (a, d’, h’) for which the proportion of retweets directed towards a during hour h’ of day d’ among all retweets of day d’ is higher than the expected proportion for this author during this hour on all other days, for all authors during a previously found abnormal hour (d’, h’).

Entity = (a, d’, h’)
Context: Coverage
Activity: v_{obs} = v_{(a, d’, h’)} / v_{(a, d’, h’)}
Global: Days × (d’, h’)
Local: V × (d’, h’)

ex: On the 24th of August from 19:00 to 20:00, Nicolas Sarkozy has been much more retweeted than he is usually during this hour when looking at other days.

Results Comparison with news records

Leading authors: – (24th, 19h): Political Meeting of Nicolas Sarkozy.
– (27th, 15h): Political Meeting of Alain Juppé.
– (24th, 20h): Interview of Nicolas Sarkozy on television news.

No leading authors: – (24th, 21h): Public reaction to Nicolas Sarkozy’s interview that took place one hour before.

References

Contact Us | E-mail: audrey.wilmet@lip6.fr | Web: www.complexnetworks.fr