Clustering Human Activity Expressions

Steve Wilson and Rada Mihalcea, University of Michigan
{steverw,mihalcea}@umich.edu

Introduction

- Our activities say a lot about who we are:
- We can't always directly observe human activities, yet people talk about what they are doing online:

Personality, Values, Interests [Ajzen 1987; Rokeach 1973; Goecks and Shavlik, 2000]

Diet and exercise [Saltonstall 1993; Maher et al. 2014]

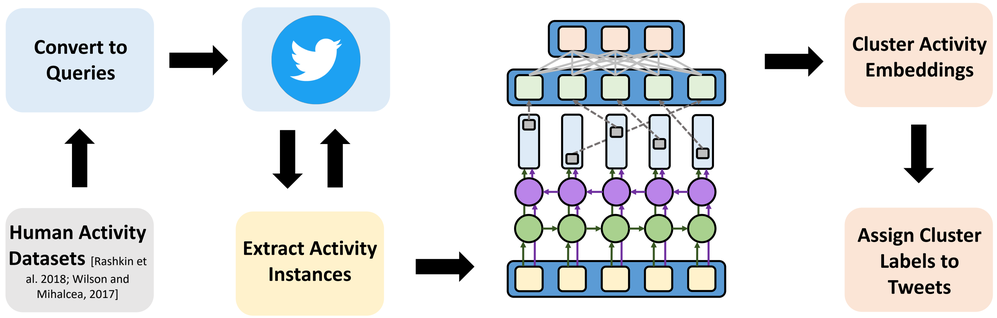
Potential buying behaviors [Zhang and Pennacchiotti, 2013]

Future Actions and Decisions [Ouellette Wood, 1998; Gerber, Green, and Shachar, 2003]

- However, it can be difficult to automatically reason about such as wide range of possible activities..
- Further, there are countless ways to express the same activity:

- **Our goal:** build a semantic space of activities into which we can map any unknown string describing an activity.

Dataset Construction



	count	unique
Event2Mind activities	24,537	24,537
Survey activities	5,000	4,957
Total	**29,537**	**29,494**

Table 1: Activities data size.

Total queries	29,494
Query results	422,607
Avg. results/query	14.33
Num. unique users	358,091
Users with profiles	96.9%
Users with addtl. tweets	95.9%

Table 2: Details of query results.

- **For each user who posted a query match, also collect:**
  - Previous tweets (up to most recent 3200)
  - Profile information
  - Other possible activity mentions

addtl. activities extracted	22,725,671
avg. addtl. activities / user	63.46
% users with $\geq$ 1 addtl. activities	93.5%
% users with $\geq$ 5 addtl. activities	87.1%

Table 3: Additional data for users in dataset.

Clustering Results

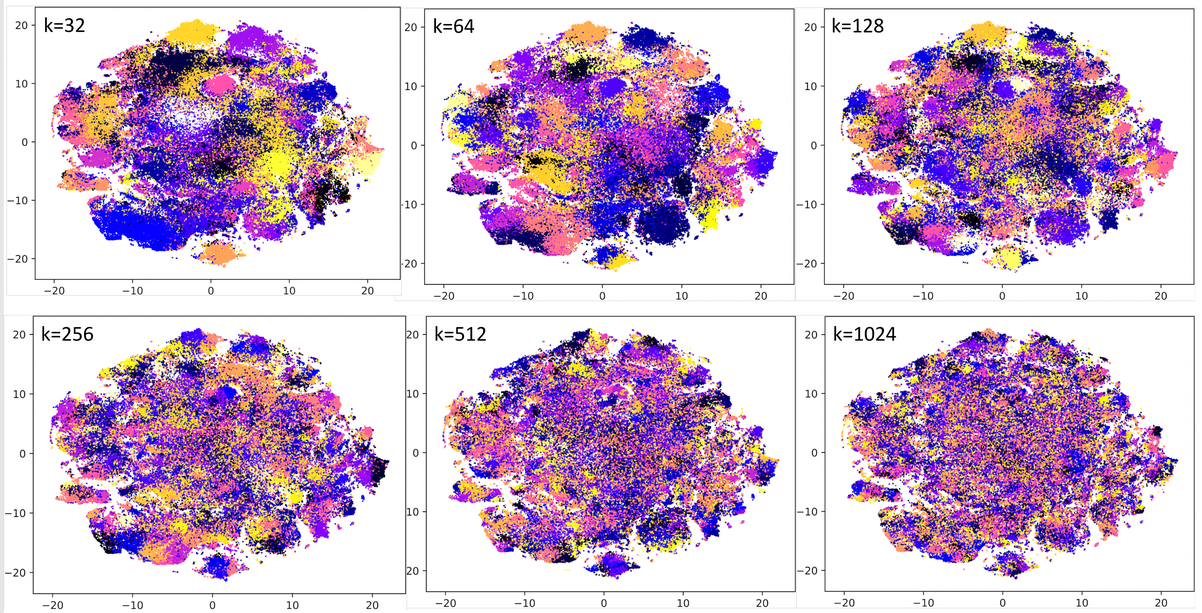


Figure 1: T-SNE projections of k-means clustering results for varying values of k.

Example Clusters

- Get a new iphone
- Buy new shoes
- Buy a toy
- Go and get some earrings
- Make a pink dress

- See a dog online
- Try to walk one's cat
- Love horses
- Find a dog wandering around the streets
- Hear a dog barking

- Take a trip to California
- Spend a year in Afghanistan
- Come of age during the Vietnam draft

- Eat a whole thing of almonds in one sitting
- Eat popcorn in bed
- Get tacos by oneself
- Stop eating fast food

- Run cross country
- Reach the top of the ladder
- Go to the doctor's with pain in one's toe
- Start a fight

- Train hard
- Run home
- Run a half marathon
- Gain weight too much
- Run out of steam

Future Directions

- Setup prediction framework to automatically determine which clusters are more likely for a given person to perform given what we know about them:
  - Previous tweets
  - Previous activities
  - Profile information
- Explore relationships between clusters and:
  - Other clusters
  - Inferred personal values

Acknowledgements

This material is based in part upon work supported by the National Science Foundation (#1344257), the John Templeton Foundation (#48503), the Michigan Institute for Data Science, and the National Institute of Health (#5R01GM112697-02). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the John Templeton Foundation, the Michigan Institute for Data Science, or the National Institute of Health.

Images:
http://www.redcross.org/images/MEDIA_CustomProductCatalog/m9040073_514x260_BAL_disaster_cambro.gif
https://marathon.runinbudapest.com/wp-content/uploads/sites/3/2015/05/maraton_02-1024x300.jpg
http://news.fiu.edu/wp-content/uploads/Father-reading-to-child.jpg
https://timedotcom.files.wordpress.com/2017/05/gettyimages-138711450.jpg