Data Analysis Tools for Informing Opioid Policy Decision – Embeddings

Rajan Vatassery, PhD

Opioids are substantially overprescribed, frequently with harmful consequences. Studies show that prolonged opioid prescriptions can increase the frequency of an opioid overdose-related emergency by 60% (Kuo 2016). This epidemic is most prevalent in poor, uninsured, urban residents of generally Caucasian or African-American origin (McDonald 2012). Minnesota in particular has a substantial population of Native Americans that are affected by the opioid epidemic as well (Minnesota Opioid Action Plan 2018)

While some policies have reduced the severity of the opioid crisis, new policies will need to be enacted to continue this progress. In order to determine the best policies for the State of Minnesota, the various factors driving this crisis need to be drawn out. Fortunately there are many data analysis tools at the disposal of researchers in this field.

One such tool is matrix factoring, which is frequently used in movie recommendation algorithms. A researcher using this technique to investigate future opioid policy can provide a matrix containing historical opioid use or prescriptions, sorted by two or more data attributes, such as year, location, distance to physician, age, ethnicity, gender, or other metrics. **Table 1**, below, shows an example of such a matrix correlating opioid use by year and county in Minnesota.

County	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AITKIN	84.5	63	99.9	118.2	114	117.8	122.9	113	107.2	98.7	80.6	68
ANOKA	53.4	57.6	61.1	66.2	67	68.3	65.3	62.4	61.6	57.3	52.1	44.9
BECKER	46.3	54.2	57.7	55.1	59.9	57.7	71.6	66.4	59.4	52.6	45.6	37.5
BELTRAMI	61.1	60	69.6	74.2	73.7	75.3	89.1	70.2	67	60.1	55.4	46.4
BENTON	48.3	51.6	54.1	53	52.6	44.7	51.1	47.5	47.4	43	39.3	33.1
BIG STONE	96.1	112.5	118.5	115.7	113.1	124.7	129.1	111.8	107.9	88.4	84.9	83.4
BLUE EARTH	67.7	71.3	73.3	70.4	69.1	70.6	78.9	74.8	72.7	70.7	66	58.5
BROWN	67.3	71.7	65	64.5	65.8	63.2	82.1	77.8	72.5	67.6	56.6	48.9
CARLTON	70	77.7	81.6	96.6	101	95.9	113.3	106	99.1	72.1	62.9	54.6
CARVER	38.6	42.2	43.7	50.2	50.5	53.2	52.1	46.7	46.8	43.7	40.2	35.6
CASS	38.7	41.8	42.1	44.8	47.4	49.1	52.1	48.1	42.2	35.9	31.2	24.7
CHIPPEWA	18.4	23	36.9	56.5	60.5	64.3	74.2	74.1	68.3	73.3	70.4	62.8
CHISAGO	30.1	32.1	30.2	26.8	28.4	32.5	29.5	36.5	39.9	41.1	34.6	30.9
CLAY	54.2	58.6	59.4	54.8	54.8	56.7	67.0	65.7	66.0	62.7	54.4	44.8

Table 1: The top few lines of a table correlating opioid prescriptions sorted by county and year

Once the data is in matrix form the user can factor the matrix into vectors for each data attribute. In the case of the current analysis the matrix is factored into a series of vectors with one vector for each county, and another series of vectors with one vector for each year the data covers, as shown in **Figure 1**. The size of the vectors is defined by the analyst before the matrix factoring. Matrix factoring takes place by a steepest gradient optimization. Once the matrix is factored into these vectors, often called latent factors, the researcher can determine the similarity between two different data attributes.
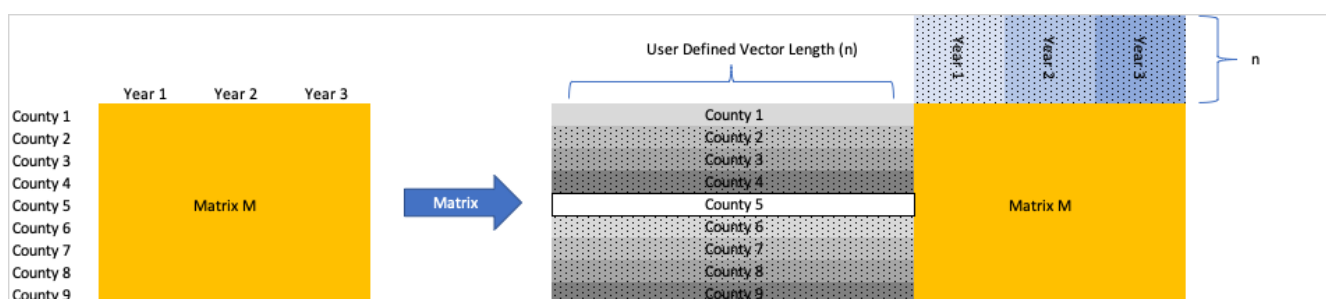
Figure 1: Matrix factoring of Matrix M into vectors for each county and each year

In the example from **Table 1** the researcher can take the set of vectors corresponding to the counties in Minnesota and use cosine similarity <u>to identify counties that are facing similar problems</u>. A higher cosine similarity indicates a larger likelihood that common factors are at work. The process described above can be found at the link below, which contains the entire code used to produce the analysis here.

https://github.com/rvatassery/MinnesotaOpioid/

Discussion

The similarities can be obvious, and the obvious similarities allow for the researcher to perform sanity checks on the analysis. For example, Hennepin and Ramsey counties are the most similar to each other in terms of their latent vectors, whereas Hennepin and Pine counties are very dissimilar. Anoka, Carver, and Dakota counties are also very similar in this analysis. (A map of Minnesota's counties is provided at the end of this paper for reference purposes).

The primary benefit from matrix factoring in the context of addressing the opioid epidemic is that the counties can be split up into groups of counties for the purposes of applying various remedies. For example, given that Anoka, Carver, and Dakota Counties are very similar, a program that produces results in one county has a substantial chance of success in the other two. Targeting counties in this way can properly direct funding dollars.

Matrix factoring as demonstrated here can provide guidance based on other data attributes as well. In the data set shown in **Table 2**, cosine similarity scores for years are output so we can assess which years the opioid epidemic saw the same set of factors operating. As expected, the most similar years are the adjacent years, but the years that are non-adjacent provide an opportunity to identify important drivers for the opioid crisis.

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
2006												
2007	0.71671902											
2008	0.65258665	0.6275747										
2009	0.7079888	0.72082974	0.89104164									
2010	0.67727013	0.77692567	0.70472908	0.6833779								
2011	0.71409242	0.79744678	0.71583729	0.69649047	0.85897338							
2012	0.83254906	0.72957718	0.66993303	0.82727042	0.6067599	0.61155414						
2013	0.71014919	0.65768231	0.7779135	0.75308792	0.71805915	0.69175694	0.76999544					
2014	0.67432239	0.6165175	0.72772719	0.77082131	0.64528066	0.68511094	0.75120348	0.90508897				
2015	0.70466668	0.64781566	0.76003869	0.73498926	0.79078228	0.66806927	0.65095945	0.84827258	0.76129703			
2016	0.80134154	0.73390386	0.66834617	0.77666391	0.67574454	0.74390434	0.8487232	0.89243887	0.86085058	0.79093048		
2017	0.80953711	0.80488334	0.60350474	0.71381284	0.71564321	0.65879882	0.70177917	0.63871516	0.6298649	0.58123736	0.71317884	

Table 2: Cosine similarity chart for years in the opioid epidemic, highlighting all correlations > 0.80

Conclusion

After a short investigation of the data a few conclusions can be drawn. Suburban counties are similar in their trends of opioid prescription, but there are also similarities between urban and suburban counties, and between suburban counties and rural counties. A number of rural counties have few similarities with other counties, meaning that the opioid issues found in those counties are probably local issues and statewide approaches would be less effective in driving down usage in those counties.

The cosine similarity matrix in **Table 2** indicates 2013 was a watershed year because most of the years from 2013 onwards are similar, but these years have little similarity with the years before 2013 until all the way back in 2007. The matrix factoring method has picked out the fact that years 2006, 2007, and then the years from 2013 onwards are similar, and this is reflected in the overall statewide trend of opioid prescription, as seen in **Figure 2**.
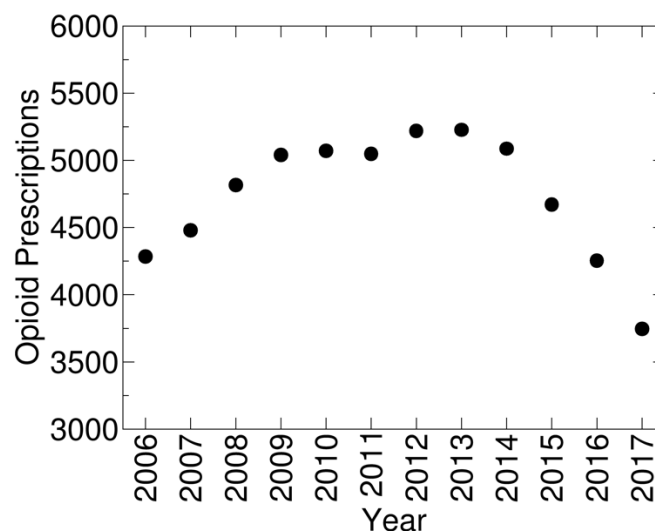
Figure 2: Overall opioid prescriptions by year across MN

Addressing the opioid epidemic is a large task to undertake, but using data science allows treatments and programs to be made to target particular populations with greater efficiencies, saving money and reaching more people with greater effect.

Kittson
Roseau
Lake Of The Woods
Marshall
Koochiching
Pennington
Beltrami
Red Lake
Polk
Clearwater
Cook
Lake
Itasca
St. Louis
Norman
Mahnomen
Hubbard
Cass
Clay
Becker
Wadena
Aitkin
Carlton
Wilkin
Otter Tail
Crow Wing
Pine
Todd
Mille Lac
Kanabec
Grant
Douglas
Morrison
Traverse
Stevens
Pope
Stearns
Benton
Isanti
Chisago
Big Stone
Sherburne
Swift
Anoka
Kandiyohi
Meeker
Wright
Chippewa
Hennepin
Ramsey
Washington
Lac Qui Parle
McLeod
Carver
Renville
Yellow Medicine
Scott
Dakota
Sibley
Lincoln
Lyon
Redwood
Nicollet
Le Sueur
Rice
Goodhue
Wabasha
Brown
Pipestone
Murry
Cottonwood
Watonwan
Blue Earth
Waseca
Steele
Dodge
Olmsted
Winona
Rock
Nobles
Jackson
Martin
Faribault
Freeborn
Mower
Fillmore
Houston

References:
Kuo, Y., *et al.*, Amer. J. Med. (**2016**) *129,* 221.e21-221.e30
McDonald, D.; Carlson, K.; Izrael, D., J. Pain. (**2012**) 13:10, 988-996
Minnesota Opioid Action Plan, 2018