Notes for future Graph Match paper

Geoffrey Iyer

January 23, 2018

NOTE TO SELF: Look up quadratic matching problems. NOTE TO SELF 2: Apparently some cool stuff could be done with brain-graphs. The vogelstein 2015 paper says that there's been a lot of failure in this area due to size of data problems. So maybe we can be the new cool kids on the block. NOTE TO SELF 3: Need to check out the QAP benchmarks

So the idea here is that I'm going to present the graph matching algorithm as an advancement purely on the computational side. Between hierarchical and Nystrom we go really fast. But first we have to go through the current state of the art. That's what this file of notes is for. And assuming that I'm correct and we are the fastest, this can be turned into a related work section for the final paper.

1 Results from my code

Just the timing: I ran graph match on sets of size $100 \cdot 2^i$ for $i = 1, \ldots, 10$. For each size I did 5 trials and took the average time (there was still a lot of variance so I probably have to do more trials). I didn't think very hard about setting the cluster size for the hierarchical part, that could probably be optimized. Here are the results.

Many-to-many matching

i	size	avg Time	ratio
1	200	0.0495	N/A
2	400	0.0375	0.7582
3	800	0.0601	1.6043
4	1600	0.1205	2.0034
5	3200	0.3294	2.7334
6	6400	0.8166	2.4794
7	12800	2.0106	2.4620
8	25600	4.9548	2.4643
9	51200	16.1501	3.2595
10	102400	55.8308	3.4570

One-to-one matching

i	size	avg Time	ratio
1	200	0.1135	N/A
2	400	0.1776	1.5657
3	800	0.4549	2.5608
4	1600	1.3227	2.9078
5	3200	3.5639	2.6943
6	6400	12.2408	3.4346
7	12800	24.4615	1.9984
8	25600	86.9463	3.5544

So in both situations it looks like we're beating n^2. The best algorithm I found so far in papers is n^3, so in terms of speed we're pretty cool. Of course we still have to worry about accuracy.

2 2010-Present

2.1 Vogelstein Fast Approximate Quadratic Programming [1]

2015, Seems pretty relevant

He's doing the problem for data of size 10^6. Seems pretty relevant... I didn't fully understand the algorithm but I saw Hungarian in there and I'm worried that he's going too slow. He definitely has big-$O(n^3)$. It says so in the paper. But he claims the leading constant is sufficiently small to handle the thingy.

In the experimental section he says he does 1000 nodes in 300 seconds (approx). So this code is in a completely different category. The good news for him is that he gets really good accuracy.

QAPLIB is a library is 137 quadratic assignment problems, but they are all small. Like 250 vertices. Not what we are focused on.

Erdos-Renyi is an algorithm for generating random graphs. Could be good for accuracy tests?

2.2 Fiori Robust Multimodal [2]

2013, SEMI-RELEVANT

Minimizing a similar energy but not the same. The most important takeaway is that they match at most graphs of size 500. So totally not on the same level. I think I can leave this be.

3 2000-2009

3.1 Kosinov Eigenspace and Clustering [3]

2002, NOT-RELEVANT

A similar idea of going to eigenspace and matching graphs there. Includes the bit about aligning eigenvectors. Doesn't talk about size of graphs.

3.2 Knossow Inexact [4]

2009, RELEVANT

The most related of related stuff. I very much use their ideas. But they max out at like 5,000 or 7,000 nodes so that's interesting.

They are more unsupervised than me though which is nice for them.

4 Not state of the art but related

4.1 Hoffman Variation of Spectrum [5]

This is a foundational paper of the graph matching idea. It relates the norm minimization problem to eigenvalues and permutations.

4.2 Umeyama [6]

You've already read this a million times. First paper with the good idea. Does some silly stuff for eigenvector alignment but background theory is good.

References

[1] JT Vogelstein, JM Conroy, V Lyzinski, LJ Podrazik, SG Kratzer, and ET Harley et al. Fast approximate quadratic programming for graph matching. *PLoS ONE*, 10(4), 2015. 2.1

[2] Marcelo Fiori, Pablo Sprechmann, Joshua T. Vogelstein, Pablo Muse, and Guillermo Sapiro. Robust multimodal graph matching: Sparse coding meets graph matching. In C.j.c. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 127–135. 2013. 2.2

[3] Serhiy Kosinov and Terry Caelli. Inexact multisubgraph matching using graph eigenspace and clustering models. In Terry Caelli, Adnan Amin, Robert P. W. Duin, Dick de Ridder, and Mohamed Kamel, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, pages 133–142, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg. 3.1

[4] David Knossow, Avinash Sharma, Diana Mateus, and Radu Horaud. *Inexact Matching of Large and Sparse Graphs Using Laplacian Eigenvectors*, pages 144–153. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. 3.2

[5] A. J. Hoffman and H. W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Math Journal*, 20(1):37–39, 1953. 4.1

[6] S. Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(5):695–703, September 1988. 4.2