Application of kernels to link analysis (Approximation methods)

Takahiko Ito

Nara Institute of Science and Technology

Background

- A huge amount of data can be represented by graphs.
  - WWW, citation or social networks
    - Node: web page, person
    - Edge: hyperlink, citation

  - We can get useful information from these types of graph data, however …

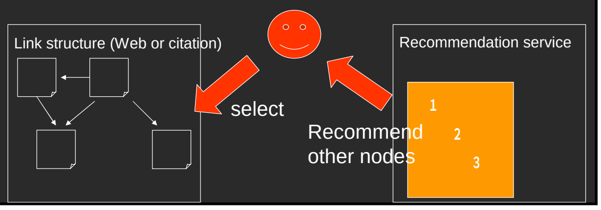
Motivation

- Exploring huge graphs is a difficult task.
  - Ex. Visualization techniques can show only a fraction of huge graphs at a time.
- Services to explore huge graphs data are desired!

Motivation (recommendation service)

1. Users select favorite nodes
2. Recommendation services suggest nodes
   - related to the selected nodes
   - popular in the graph

   with rank



Previous work

- We have shown graph kernels compute relatedness or relative importance between two nodes.
  - → They are adaptable for recommendation services.

Problem of graph kernels

- Graph kernels are computationally inefficient $O(|N|^3)$
  - Where |N| is the number of nodes.
- → We propose two types of approximation methods for graph kernels

Table of contents

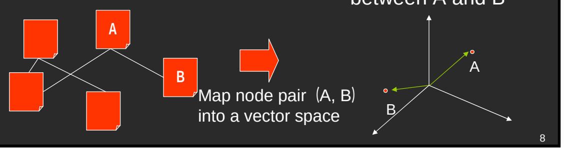
- Review graph kernels
  - Neumann kenels [Kandola et al., 2003]
  - Regularized Laplacian kernels [Smola and Kondor, 2003]
- Propose two types of approximation methods for graph kernels

Graph kernels

- define an inner product of nodes in a graph.
  - Heat kernels [Chung, 1997; Kondor & Lafferty, 2002]
  - Neumann kernels [Kandola et al., 2003]
  - Regularized Laplacian kernels [Smola & Kondor, 2003]
  - …

Ex:

Compute the inner product between A and B



Computation of graph kernels

- Graph kernels are represented by weighted sum of matrices.

Computation of graph kernels (Neumann kernels)

- Neumann kernels [Kandola et al., 2003] :

  $NK_{\beta}(A^TA) = A^TA + \beta(A^TA)^2 + \beta^2(A^TA)^3 + \ldots$

  where
  - A: adjacency matrix of graph G
  - β: diffusion rate

Computation of graph kernels (Regularized Laplacian kernels)

Regularized Laplacian kernel matrix [Smola and Kondor, 2003]

$RLK_{\beta}(S) = I + \beta(-L(S)) + \beta^2(-L(S))^2 + \beta^3(-L(S))^3 + \ldots$

where
- S: symmetric matrix (such as A^TA)
- L(S): Laplacian [Chung, 1997] of S

  $L(S) = D(S) - S$
- D(S): Diagonal matrix
  - (i,i)-element represents the degree of node i in the graph induced by S

Table of contents

- Review graph kernels
- Propose two types of approximation methods for kernels
  - Limited steps approximation
  - Limited eigenvalues approximation

Approximation by limited steps 1/2

Standard techniques for matrix computation allow the approximation of kernel computation with the sum of the first k terms of infinite series.

Limited steps approximation of Neumann kernels:

$NK'_\beta(A^TA) = A^TA + \beta(A^TA)^2 + \ldots \beta^{k-1}(A^TA)^k$

The approximation error is bounded:
$(|N|/K!)(r\lambda)^{-1}-1)^{-1/2}$

13

Approximation by limited steps 2/2

- If one is concerned with the importance of node relative to a single node rather than entire kernel matrix,
- ➔ We can reduce space requirement by limited step approximation:
  i-th column of approximated Neumann kernel:
  $(A^TA)x_i + \beta\,(A^TA)^2x_i + \ldots \beta\,(A^TA)^kx_i$

where x_i: if k = i $x_i(k) = 1$; otherwise $x_i(k) = 0$

14

Approximation by limited eigenvalues 1/2

- Neumann kernels can be represented as:
  - $NK_\beta(A^TA) = A^TA + \beta(A^TA)^2 + \beta^2(A^TA)^3 + \ldots$
    $= P^TDP + P^T\beta D^2P + \ldots$
    $= P^T(\Sigma\, \beta^kD^{k+1})\, P$

A^TA is decomposed as P^TDP

where P consists of eigenvectors of A^TA,
D contains of eigenvalues of A^TA.

15

Approximation by limited eigenvalues 2/2

- Use only the k largest eigenvalues in A^TA
  - N.B. Use D' in place of D
    D': if $(D(i, i) > \lambda^k)$ $D'(i, i) = D(i, i)$; Else $D'(i, i) = 0$
    λ^k = k-th largest eigen value
- Definition of Neumann kernels with limited eigenvalue approximation:
  - $NK'_\beta(A^TA) = P^T(\Sigma\, \beta^kD'^{k+1})\, P$

16

Experiments (performances of approximation methods)

Compare
- Graph kernels (Neumann kernels, regularized Laplacian kernels)

and
- Limited step approximation, limited eigenvalues approximation

dataset:
citation graph (2687 papers on natural language processing)

17

Experimental settings

- Process of experiments:
  1. Extract the row vector for each paper from the kernel matrices and rank papers based on the magnitude of elements.
  2. Compare the ranking of each node among kernels using k-min distance.
- K-min distance:
  - If two lists have similar rankings ⇒ K-min distance is small
  - If two lists have similar rankings ⇒ K-min distance is large

18

Results (limited step approximation)

Averages of K-min distances between kernels and their limited step approximations

# of steps		5	10	50	100
Neumann	$\beta = 0.9$ * spectral radius^{-1}	10.8	6.7	0.7	0.5
	0.1	0.7	0.6	0.6	0.6
Regularized Laplacian	0.9	3.7	3.7	3.3	3.2
	0.1	3.1	3.1	3.1	3.1

Results (limited eigenvalues approximation)

Averages of K-min distances between kernels and their limited eigenvalue approximations

# of eigen vectors		10	50	100	500	1000
Neumann	$\beta = 0.9$ * spectral radius^{-1}	52.9	36.0	24.4	7.12	0.69
	0.1	63.9	39.5	28.4	17.4	0.73
Regularized Laplacian	0.9	76.1	70.7	67.3	50.7	31.5
	0.1	75.8	70.9	67.5	51.0	32.0

Conclusions

- We have proposed two types of approximation methods for graph kernels.
  - Examine the performances of approximation methods on citation network.

Other contributions

- Parameter estimation methods for kernels
- Some proofs
  - Non-negativity of kernels
  - Positive semi-definiteness

Future work

- Comparison between other kernel methods and traditional relatedness measures.
- Application to collaborative filtering or relevance feedback.

References

- F. R. K. Chung. *Spectral Graph Theory*. AMS, 1997
- R. Fagin, R. Kumar, D. Sivakumar. Comparing top *k* lists. In *Proc. SODA*, 2003.
- M. M. Kessler. Bibliographic coupling between scientfic papers. *American Documentation* 14:10-25, 1963
- J. Kandola, J. Shawe-Taylor, and N. Cristianini. Learning semantic similarity. In *Proc. NIPS* 15, 2003.
- R. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *Proc. ICML*, 2002.
- H. Small. Co-citation in the scientific literature: a new meature of the relationship between two documents. *J. Am. Soc. Information Science*, 24:265-269, 1973.
- A. J. Smola and R. Kondor. Kernels and regularization of graphs. In *Proc. 16th COLT*, 2003.
- S. White and P. Smyth. Algorithms for estimating relative importance in networks. In *Proc. ACM SIGKDD*, 2003.