Computing Citation Relatedness Using Kernels

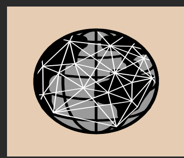
Takahiko Ito

Nara Institute of Science and Technology

1

Graph data are ubiquitous

- 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 …

2

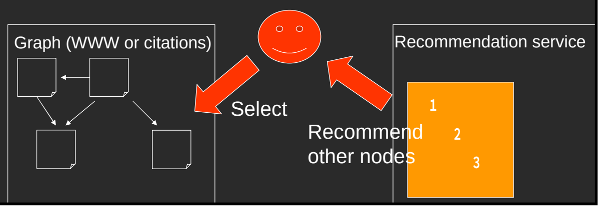
Motivation

- 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 desirable!

3

Recommendation service for graph data

1. Users select favorite nodes (root nodes) – papers / web pages
2. Based on links or citations around the root nodes, the service recommends other nodes that may have some relatedness to the root nodes.



To recommend nodes

- Relatedness measures:
  - Measures for analyzing the relationship among nodes in graphs based on graph structures.
  - However, classical relatedness measures have some limitations, if they are applied to recommendation services.
- We propose to extend traditional relatedness measures based on kernel methods.

5

Outline

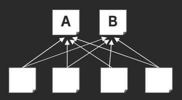
1. Introduce traditional relatedness measures
2. Two problems with traditional relatedness measures
3. To overcome the problems, we apply two kernel methods as relatedness measures.
   1. Neumann kernel [Kandola et al., 2003]
   2. Regularized Laplacian kernel [Smola and Kondor, 2003]
4. Experiments

6

Co-citation/bibliographic coupling "relatedness"

Co-citation coupling [Small et al., 1973]

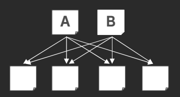
defines relatedness as the number of papers jointly citing the given pair of papers



Co-citation coupling (A,B) = 4

Bibliographic coupling [Kessler,1963]

defines relatedness as the number of common citations made by two papers



Bibliographic coupling (A,B) = 4

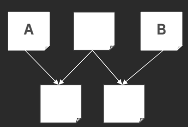
Computing co-citation/bibliographic coupling

Given adjacency matrix A of a citation graph,

- (i, j)-element of A^TA
  - ➔ Co-citation relatedness between nodes i and j
- (i, j)-element of AA^T
  - ➔ Bibliographic relatedness between nodes i and j

Problem with classic relatedness 1

- If a pair of papers does not jointly cite or is not jointly cited by any paper, co-citation and bibliographic coupling cannot measure relatedness between the two nodes.



bibliographic coupling (A,B) = 0

Problem with classical relatedness 2

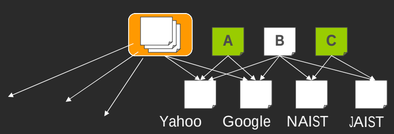
- Intuition behind bibliographic coupling relatedness:

  Two papers are related if they jointly make citation to one or more papers.

- But the number of *other* citations to the cited papers are ignored.

Problem with classic relatedness 2: Illustration

Which of A or C is more related to B?



Intuition:

C is more related to B than A is, because A and B only share citations to "generic" (or "popular", or "authoritative") pages (Google and Yahoo).

Neumann kernels [Kandola et al., 2003]

- Original Neumann kernels compute document relatedness, but *not* on the basis of citations.
- They use graphs induced from the content of documents:

  An edge between nodes (documents) has a weight based on the number of common terms in their contents.

Definition:

$NK_\beta(XX^T) = XX^T + \beta(XX^T)^2 + \beta^2(XX^T)^3 + \dots$ (document relatedness)

$NK_\beta(X^TX) = X^TX + \beta(X^TX)^2 + \beta^2(X^TX)^3 + \dots$ (term relatedness)

where X is a document-by-term matrix, and β is a "diffusion rate" parameter.

Neumann kernels for citation analysis

Neumann kernels in this work
- are applied directly to citation graphs.
- i.e., use adjacency matrix A of a citation graph in place of document-by-term matrix X.

Definition:

$NK_\beta(AA^T) = AA^T + \beta(AA^T)^2 + \beta^2(AA^T)^3 + \ldots$

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

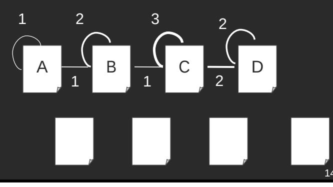
What do $(AA^T)^n$ and $(A^TA)^n$ in these series represent?

13

Meaning of $(AA^T)^n$

- Element (i, j) of $(AA^T)^n$ = number of paths of length n between nodes i and j in a bibliographic graph.
  - Where bibliographic graph is derived from AA^T
- Example:

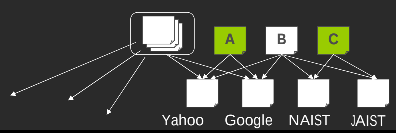
Bibliographic graph (derived from AA^T)



14

Does Neumann kernel solve the two problems?

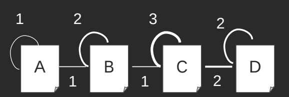
- Neumann kernels (with non-zero diffusion rate β) can give a value to a pair of nodes as long as there is more than one path between them in the bibliographic graph. Thus it does not suffer from Problem 1.
- However, Neumann kernels can not solve problem 2; they mistakenly regards A as more related to B!



15

Why Neumann kernels does not solve problem 2

- Neumann kernels compute the weighted sum of $(AA^T)^n$ with n = 1~∞



$$(AA^T)^5 = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 146 & 403 & 753 & 496 \\ 403 & 1302 & 2901 & 2002 \\ 753 & 2901 & 7454 & 5306 \\ 496 & 2003 & 5306 & 3800 \end{pmatrix} \end{array}$$

- At n=1, $(AA^T)^n$ represents the bibliographic matrix
  As n is increased…
- After n=5, all rows of $(AA^T)^n$ give an identical ranking C>D>B>A. This ranking is not relatedness among nodes but the HITS hub ranking.

16

HITS [Kleinberg, 1999]

- computes "importance" of each node
- assigns two scores to each node:

  Authority score：

  Nodes cited by many nodes receive a high authority score

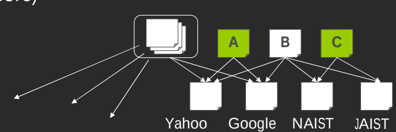
  Hub score:

  Node citing many authoritative nodes receive a high hub score.

17

Summary of Neumann kernels

- Neumann kernels are not a relatedness measure because they bias towards importance.
  Ex. Neumann kernels give a larger value to A than C with respect to B (importance (A) > importance (C) in HITS hub score)



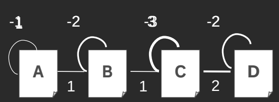
- We need to prevent Neumann kernels from biasing toward importance

18

Solution to importance bias problem

- Change weights assigned to self-loops
  - negative of the number of non-loop edges at each node
- Compute sum of weights of all paths between the nodes (unchanged from Neumann kernels)

→ Nodes with a large number of edges (important nodes) receive a large discount



Graph induced by –L(AAT)

19

Regularized Laplacian kernels
[Smola and Kondor, 2003]

Regularized Laplacian kernel matrix

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

where

- S: symmetric matrix (such as A^TA or AA^T)
- 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

20

Experiments

Compare

- Regularized Laplacian kernels

with

- Co-citation coupling

Dataset:

Citation graph consisting of 2687 papers on natural language processing

21

Regularized Laplacian kernel vs. co-citation coupling

Top ranked papers with respect to:
Marilyn A. Walker and Johanna D. Moore. Empirical studies in discourse. *Computational Linguistics Vol. 23, No. 1, 1997*

RLK	Co-cite	Title
1	1	Empirical studies in discourse
2	1	Effect of … computer spoken natural language dialogue
3	1	Message Understanding Conference tests of discourse …
4	1	The reliability of a dialogue structure coding scheme
5	1	Assessing agreement on classification tasks: …
6	1	Attention, intentions, and the structure of discourse
7	1	Building a large annotated corpus of english: the Penn Treebank
8	n/a	A prosodic analysis of discourse segments in …
9	n/a	Centering: a framework for modeling the … discourse
10	n/a	Combining multiple knowledge sources for discourse …

22

Conclusions

- Two types of kernel methods (Neumann kernel and regularized Laplacian kernel) have applied to solve the problems in traditional relatedness measures.
  - The two limitations in co-citation and bibliographic coupling relatedness can be overcome using the regularized Laplacian kernels.

23

Future work

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

24

References

- M. M. Kessler. Bibliographic coupling between scientific papers. *American Documentation* 14:10-25, 1963
- J. Kandola, J. Shawe-Taylor, and N. Cristianini. Learning semantic similarity. In *Proc. NIPS* 15, 2003.
- A. J. Smola and R. Kondor. Kernels and regularization of graphs. In *Proc. 16th COLT*, 2003.
- H. Small. Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Am. Soc. Information Science*, 24:265-269, 1973.
- S. White and P. Smyth. Algorithms for estimating relative importance in networks. In *Proc. ACM SIGKDD*, 2003.