IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 4, Issue 11, November 2015*

# Novel Approach for Policy Network Extraction from Web Documents

**Miss. Archana Loni[1], Mrs. Aarti Waghmare[2]**

Student Computer, DYPSOET, Pune, India[1]

Professor, Computer, DYPSOET, Pune, India[2]

**Abstract:** The increment of information in the web is too large, so search engine come to play an important role to find relation between input keywords. Policy networks are one well defined domain for research. In Today's World Policy networks are used by economists and political scientists. The analysis of policy networks demands a series of arduous and time-consuming some steps which are manual including interviews and questionnaires .We are calculating the strength of relation between actors in policy networks using feature extracted from data harvested from the web. Features are like out links, webpage counts. Features are like out links, webpage counts, and derived some lexical information from web snippets, documents. The features are evaluated both in jointly and isolation for both positive and negative actor relations Performance is measuring in terms of co-relation between the human rated and the automatically extracted relations.

**Keywords:** Policy Network, relatedness metrics, web link, web Documents.

## I. INTRODUCTION

The Today's governance reflects a shift away from the traditional notions of hierarchy of various forms of public policy making which are like more cooperative. Within this context, the political, social, and economic spheres to which the set of various types of people are related to explain the term "network". "Policy network" is defined as "a set of people in each of which having curiosity, or "stake" in a given policy sector and the capacity to help for find out the failure or success of particular policy. Policy networks are one well defined domain for research. For analysis of social and financial occurrence Political scientists use policy networks, especially, Political scientists also focus on growth of relations between people who are related to particular policy and the effectiveness of policies toward the formation of partnerships among actors. A policy network can be described by its actors, their linkages, and its border. Policy networks includes a cluster of public as well as private actors and a number of linkages between them that perform as media for communication and the exchange of information, expertise, trust, and other policy resources. The functional relevance and structural embedded are also use to determine network boundaries which are not initially by formal institutions.

There are some experts who perform some manual procedure for identification of policy networks. Identifying actors, links, and boundaries, i.e., manually we can collect data through taking interviews and asking questions and also requires some polished technologies for analyzing a policy network. In manual identification of the policy network subjective factors of human are strongly realize ,the manual identification is the procedure which realize strongly on the subjects of human, these subjects of human are participates in the interviews. These type of subject factors are involve the opinion, willingness to take participate and cultural issue of the person.

For collection of data or network analysis when resources are limited or unavailable then political scientist are return by using their knowledge to construct a network topology or by analyzing quantitatively, significantly limiting the evidence-based validation of their results. We have presented different methods to overcome issues related efficient automatic extraction of policy networks. Policy network use some features which are derived from the data cutting from the web for estimate the strength of relations between actors. The strength of link between two actors which represent in the policy network is computed or determined by using three types of features on documents or part of document which is search or downloaded by the various search engines.

For estimate the link strength of each pair of actors which are present in policy network for each type of feature and for their combinations a types of similarity metrics are used. The proposed method aims to be efficient and reduce human biases. We can also assume a policy network as a one type of graph in which actors are present which acts as a nodes present in given policy field and relation between that actors are represent as a edges of that network.

## II. LITERATURE SURVEY

The data which is large amounts computationally analyze by political analysts has flourished in the past few decades facilitating the study of group connections. In the literature we have presented different methods and approaches those are used for political dataset analysis. Social network analysis as well as text analysis is some of computational methods which are widely used in political science. More specifically, political analysts have used text mining to analyze electoral campaigns, verification of voters" profiles, determine positions which is ideological, code political interaction, and detect political conflicts content [3], [4]. For extraction of economic and social policy The

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

system is proposed which is known as WORDSCORES that dimensions based on word frequencies from manifestos in[5][6].
The political public statements are most of present in the various textual data rather that this copied or same speeches and political statements are also used. Similarly, the WORDFISH system mines policy dimensions of parties in [7]. The political scientist is searching the various opinions of experts by copied speeches and blogs or texts in [8]. Political scientists work on the some active areas like opinion mining which is nothing but asking questions to the experts which we can say is primary method for active research for political scientist.
Important research questions include the selection of words and terms, the scores assigned to each term, as well as, for the combination of evidence the computational model used e.g. in [9].There are various types of political sentiments which are got from the different blogs combined by lexical feature during the 2008 US Presidential election in [10]. There are some important steps or stages like the social networks in the procedure of extraction which are identification of relation in [12], also determine the relation of the two actors who are present in the same policy network in [16][17],evaluates the power of the relation of two actors present in the policy network in [18], here in this verify that the relation which is in between two actors is good or bad. There are some advance techniques are used like contacts through E-mail for creating both type of relationship in network like professional and personal in[21]. Social networks are also extracted in other types well as updated from various news articles in [23]. By using the features which are taken from the cutting data of internet the authors are gouge or examine the relations between two actors who are present in policy network in [1]. The proposed method presented in [1] was automatic and there is no any requirement of outer way of knowledge rather than any word which specify the political actor. The features are valuable for both the cases for single or for combination of opposite (i.e. positive and negative) relations of actors.

Problem definition: In the policy network the actors are act like nodes and their relations are like edges of the network. The relation of two actors is not only cooperative but it also can be opposed to each other. Often policy networks are deeply observed at their childhood when the links or edges between actors are not directly just come to view and may be examine through direct communication. A variety of features got from various documents or part of that documents which are observed actors relation in policy network.

### III. PROPOSED FRAMEWORK

The system architecture of proposed system is shown in following fig.

Relatedness metrics: There are three types of the relatedness metrics that are 1) Page count base metrics. 2) Text based metrics. 3) Link based metrics. Mathematical presentation of each is as follows. And the mathematical representation of each is as follows.

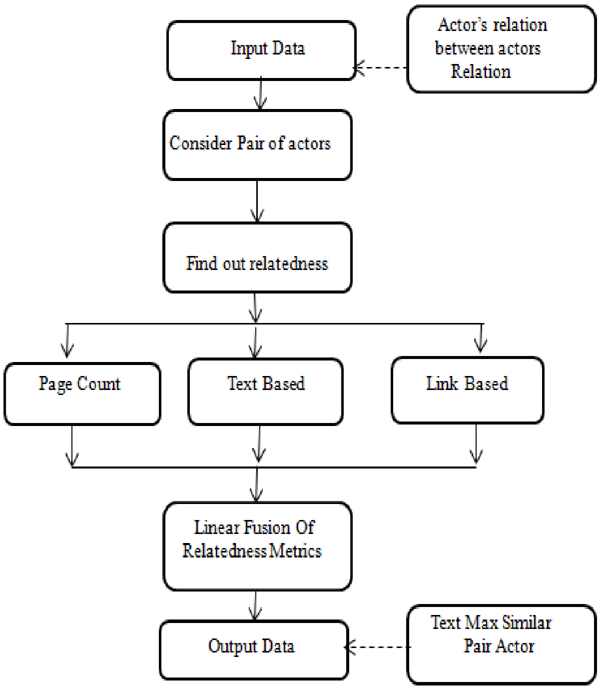


Fig.1. System Architecture

A. Page Count Base Metrics
In this type of metrics first of all consider set of documents which are recorded in the form of index. Consider {D} as set of document and also cardinality is denoted as |D|. These all the documents are searched from search engine. The documents which are arranged by $a_i$ and that is denoted by $\{D_{a_i}\}$.Similarly consider the document in which both the actors $a_i$ and $a_j$ are present which is denoted by $\{D_{a_i,a_j}\}$ and cardinality is denoted as $|D_{a_i,a_j}|$. Again in Page Count metrics is divided in to some types that are as follows.
Jaccard Coefficient: The Jaccard Coefficient is denoted by $S_J^P$ and equation is as follows,

$$S_J^P\ (a_i, a_j)=\frac{|D_{a_i,a_j}|}{|D_{a_i}|+|D_{a_j}|-|D_{a_i,a_j}|}\ . \qquad 1$$

Dice Coefficient: The Dice Coefficient is denoted as $S_D^P$ and equation is as follows, $S_J^P$

$$S_J^P\ (a_i, a_j) = \frac{2|D_{a_i,a_j}|}{|D_{a_i}|+|D_{a_j}|} \qquad 2$$

Google-based semantic relatedness**:** The Google-based semantic relatedness is denoted as $S_R^P$ and the mathematical equation is as follows,

$$S_R^P(a_i, a_j)=\frac{\max\{\log|D_{a_i}|,\log|D_{a_j}|\}-\log|D_{a_i,a_j}|}{\log|D|-\min\{\log|D_{a_i}|,\log|D_{a_j}|\}} \qquad 3$$

B. Text Based Metrics:

$$S_W^T(a_i, a_j)=\frac{\sum_{l=1}^{N} v_{a_i,l} v_{a_j,l}}{\sqrt{\sum_{l=1}^{N}(v_{a_i,l})^2}\ \sqrt{\sum_{l=1}^{N}(v_{a_j,l})^2}} \qquad 4$$

The equation 4 is Text Based Metrics equation in which W is length of context window and the vocabulary size is


DOI 10.17148/IJARCCE.2015.411115

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 4, Issue 11, November 2015*

denoted by N. In this type of metrics the strength of relation is find out by observation of lexical context in documents which are present on web.

C. Link Based Metrics:
In this type of metric describes the relatedness metric for calculating the degree of association between two actors. The documents which are downloaded from web are exploits by these links. The equation of Link Based Metrics is as follows,

$$S_R^L(a_i, a_j) = \frac{\max\{\log|O_{a_i}|, \log|O_{a_j}|\} - \log|O_{a_i,a_j}|}{\log|O| - \min\{\log|O_{a_i}|, \log|O_{a_j}|\}} \qquad 5$$

Where, $\{O_{a_i}\}$, $\{O_{a_j}\}$ and $\{O_{a_i,a_j}\}$ are the set of outlinks which are for actors $a_i$ , $a_j$ and both respectively.

D. Linear Fusion of Relatedness Metrics:
Here in this metrics all the metrics of relatedness are combined for compose the score of relatedness S between two actors$a_i$, $a_j$. And that is defined as follows,

$$S(a_i , a_j) = \lambda_p S^P(a_i , a_j) + \lambda_T S^T(a_i , a_j) + \lambda_L S^L(a_i , a_j) \qquad 6$$

In above equation 6 Page Count is by$S^P$, Text Based by $S^T$, and Link Based by $S^L$ denoted. And the corresponding weights are denoted by $\lambda_p$ , $\lambda_T$, and $\lambda_L$ respectively.

## IV. CONCLUSION

In this work computes the relatedness between two actors who are present in the same policy network automatically. And for computation of relatedness here various features are also described like Page Count Metrics, Text Based Metrics and Link Based Metrics and also combination of all these features to get final relatedness between actors of policy network.

## REFERENCES

1. Theodosis Moschopoulos, Elias Iosif, Alexandros Potamianos, Senior Member, IEEE, and Shrikanth Narayanan, Fellow, IEEE, “Toward the Automatic Extraction of Policy Networks Using Web Links and Documents”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 10, OCTOBER 2013.
2. V. Schneider and P. Kenis, Policy Networks and Policy Analysis: Scrutinizing a New Analytical Toolbox, pp. 25-59, Westview Press, 1991.
3. L. Zhu, Computational Political Science Literature Survey, http://www.personal.psu.edu/luz113, 2013.
4. B. Monroe and P. Schrodt, “Introduction to the Special Issue: The Statistical Analysis of Political Text,” vol. 16, no. 4, pp. 351-355, 2008.
5. J. Garry, M. Laver, and K. Benoit, “Extracting Policy Positions from Political Texts Using Words as Data,” Am. Political Science Rev., vol. 97, no. 2, pp. 311-331, 2003.
6. K. Benoit and M. Laver, “Estimating Irish Party Policy Positions Using Computer Word scoring: The Elections 2002.
7. J.B. Slapping and S.O. Proksch, “A Scaling Model for Estimating Time-Series Party Positions from Texts,” Am. J. Political Science, vol. 52, no. 3, pp. 705-722, 2008.
8. M. Thomas and L. Lee, “Get out the Vote: Determining Support or Opposition from Congressional. Floor-debate Transcripts,” Proc. Conf. Empirical Methods in Natural Language Processing, pp. 327-335, 2006.
9. B. Chen, D. Kifer, and D. Lee, “What Is an Opinion About? Exploring Political Standpoints Using Opinion Scoring Model,” Proc. 24th AAAI Conf. on Artificial Intelligence, pp. 1007-1012, 2010.
10. W. Gryc as well as K. Moilanen, “Leveraging Textual Sentiment Analysis with Social Network Modelling: Sentiment Analysis of Political Blogs in the 2008 U.S Presidential Election,” Proc. Text to Political Positions‟ Workshop, 2010.
11. B. Monroe, M. Colaresi, and K. Quinn, “Fightin‟ Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict,” Political Analysis, vol. 16, no. 4, pp. 372-403, 2008.
12. H. Kautz, and M. Shah, “The Hidden Web,” AI Magazine, vol. 18, no. 2, pp. 27-36, 1997.
13. M. Hamasaki, J. Mori, and Y. Matsuo “POLYPHONET: An Advanced Social Network Extraction System from the Web,” Proc. 15th Int‟l World Wide Web Conf., pp. 397-406, 2006.
14. H. Tomobe, Y. Matsuo “Social Network Extraction of Conf. Participants,” Proc. 12th Int‟l World Wide Web Conf., 2003.
15. Y. Jin, Y. Matsuo, and M. Ishizuka, “Extracting Social Networks among Various Entities on the Web,” Proc. European Conf. The Semantic Web: Research and Applications, pp. 251-266, 2007.
16. J. Mori, T. Tsujishita, Y. Matsuo, and M. Ishizuka, “Extracting Relations in Social Networks from the Web Using Similarity Between Collective Contexts,” Proc. Fifth Int‟l Semantic Web Conf., pp. 487-500, 2006.
17. Y. Merhav, and D. Barbosa, “Extracting Information Networks from the Blogosphere: State-of-the-Art and Challenges,” Proc. Fourth Int‟l Conf. Weblogs and Social Media, Data Challenge Workshop, 2010.
18. R. Xiang, J. Neville, and M. Rogati, “Modeling Relationship Strength in Online Social Networks,” Proc. 19th Int‟l World Wide Web Conf., pp. 981-990, 2010.
19. P. Mika, “Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks,” J. Web Semantics, vol. 3, no. 2, pp. 211-223, 2005.
20. P. Mika, “Ontlogies Are Us: A Unified Model of Social Networks and Semantics,” J. Web Semantics, vol. 5, no. 1, pp. 5-15, 2007.
21. A. Culotta, R. Bekkerman, and A. Mccallum, “Extracting Social Networks and Contact Information from Email and the Web,” Proc. First Conf. Email and Anti-Spam, 2004.
22. A. Gruzd and C. Haythornthwaite, “Automated Discovery and Analysis of Social Networks from Threaded Discussions,” Proc. Int‟l Network of Social Network Analysis, 2008.
23. B. Pouliquen, R. Steinberg, and J. Belyaeva, “Multilingual Multi-Document Continuously-Updated Social Networks,” Proc. Int‟l Conf. in Recent Advances