International Journal of Emerging Trend in Engineering and Basic Sciences (IJEEBS)
ISSN (Online) 2349-6967
Volume 2 , Issue 1(Jan-Feb 2015), PP301-305

Alternative Method for Extraction of Policy Networks Using WWW

Mrs.Vaishali Chaudhari[1], Prof. Ratanraj Kumar[2]
[1]*(Pune University, India,* vaishalip89@gmail.com*)*
[2]*(Pune University, India)*

Abstract :- The growth of information in the web is too large, so search engine come to play a more critical role to find relation between input keywords**.** The aim of this paper is to review the state of the art in the field of policy networks and to explore their usefulness in studying metrics. It is argued that policy networks are more than an analytical tool box for studying these phenomena. The policy networks demands a series of arduous and time - consuming manual steps 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 is like outlinks, webpage counts, and lexical information extracted from web documents or web snippets. 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 networks, relatedness metrics, similarity metrics, web search.

I. INTRODUCTION

The Today's governance reflects a shift away from the traditional notions of hierarchy toward more cooperative forms of public policy making. Within this context, the term "network" is often used to describe clusters of different types of actors, who are related in the political, social, and economic spheres. term "policy network" is defined as "a cluster of actors, each of which has an interest, or "stake" in a given policy sector and the capacity to help determine policy success or failure." Political scientists use policy networks to investigate social and financial phenomena, especially, the evolution of relations between actor and the effectiveness of policy toward the formation of partnerships among actors. Policy network can be described by their linkages, its actors, and its boundary. Policy network analysis starts with three basic assumptions. First (again), modern governance is frequently non-hierarchical. Few policy solutions are simply imposed by public authorities. Governance involves mutuality and interdependence between public and nonpublic actors, as well as between different kinds of public actor. Second, the policy process must be disaggregated to be understood because 'relationships between groups and government vary between policy areas' (Rhodes 1997: 32). In other words, it makes little sense to talk generally of a 'strong state' or 'corporatist state' – let alone a 'strong' or weak' international organisation (IO) – because states and IOs are much stronger *vis-à-vis* affected interests in some policy sectors than in others.Third and finally, governments remain ultimately responsible for governance, but that is not the whole story. Before policies are 'set' by elected political actors, policy choices are shape and refined in bargaining between a diverse range of actors, including some who are nongovernmental, all of whom have an 'strategies that generate new political and economic forces' (Thatcher 1998: 406). Sometimes, they can go so far as to 'play a role in the determination of their own environment, with repercussions for the fit between political interests, organizational structures and economic objectives' (Thatcher 1998: 406; see also Dunn and Perl 1994; Peterson 1995b).

We propose an automatic method to measure semantic similarity between words or entities using Web search engines. Because of the vastly numerous documents and the high growth rate of the Web, it is not feasible to analyze each document separately and directly. Web search engines provide an efficient interface to this vast information. Page counts and snippets are two useful information sources provided by most Web search engines. Page count of a query is the number of pages that contain the query words. Snippets, a brief window of text extracted by a search engine around the query term in a document, provide useful information regarding the local context of the query term. Semantic similarity measures defined over snippets have been used in query expansion [42], personal name disambiguation [4] and community mining [6]. Processing snippets is also efficient as it obviates the trouble of downloading web pages, which might be time consuming depending on the size of the pages. However, a widely acknowledged drawback of using snippets is that, because of the huge

scale of the web and the large number of documents in the result set, only those snippets for the top-ranking results for a query can be processed efficiently.

Typically, policy networks are identified through a manual procedure performed by experts. Identifying actors, links, and boundaries, i.e., analyzing a policy network's structure, requires refined techniques and extensive and time-consuming manual collection of data through interviews and questionnaires. During the manual identification of networks, many subjective factors may be exist, because this procedure relies strongly on the human subjects that participate in the interviews. Such factors include personal opinions, the person's willingness to participate, and even cultural issues. Overall, policy network identification currently requires a "large scale investment" that does not always "lead to breathtaking empirical and theoretical results"[2]. When lacking the resources for data collection and network analysis, political scientists often revert to qualitative construct or analysis the network topology using their intuition, significantly limiting the evidence-based validation of their results. .

II. LITERATURE SURVEY

The use of computational analysis of large amounts of data by political analysts has flourished in the past few decades facilitating the study of group connections. Two computational methods have been widely used in political science namely text analysis and social network analysis. More specifically, political analysts have used text mining to analyze electoral campaigns, identify voters' profiles, determine ideological positions, code political interaction, and detect political conflict's content [3], [4]. Textual data mostly consist of political manifesto, but transcribed speeches and political statements are also used. In [5], [6], the WORDSCORES system is proposed that extracts economic and social policy dimensions based on word frequencies from manifestos. Similarly, the WORDFISH system [7] mines policy dimensions of parties and estimates their uncertainty over time using word frequencies from manifestos. opinion mining is an active research area that is also relevant to political scientist. Opinions can be mined from blogs, text or from transcribed speech, e.g., [8]. Important research questions include the selection of lexical features (words and terms), the scores assign to each of term, as well as, the computational model used to combine the evidence, e.g., [9]. In [10], lexical features are combined with social information extracted from blog to classify political sentiments during the 2008 US Presidential election. In [11], opinion mining techniques (including lexical feature selection) are applied to the analysis of political conflicts.

Regarding political analysis , social network analysis, have used network analysis to study formal and informal interactions. Policy network extraction can be considered as a special type of social network extraction, an active research area. The major step in the extraction of social network are relation identification i.e., to identify whether two actors are related, relation labelling assign an existing relation to a category and the estimation of strength [18], i.e., identify whether an existing relation is strong or weak. The most common feature used to identify a relation is the frequency of co-occurrence of the related pair of terms in web documents, but other features, such as, lexical context, keyphrases, log files, and e-mail information are also used. In [12], a network of experts with respect to certain topics is constructed by estimating similarity of users according to the frequency of concurrence of their names in web documents. Similarly, in [19], [20], web co-occurrence of entities is used for creating a network of research communities. In [13], [14], [15], web concurrences are used for the extraction of social network of conference participants, a machine learning approach is used to classify every relation from a predefined set of relation types. In [16], automatically extracted key phrases are used to describe the relations between entities. contacts are used as features in to create personal and professional relationship networks. In, social networks are extracted and updated over time using monolingual or multilingual news from articles.

III. PROPOSED APPROACH FRAMEWORK AND DESIGN

a. ARCHITECTURE

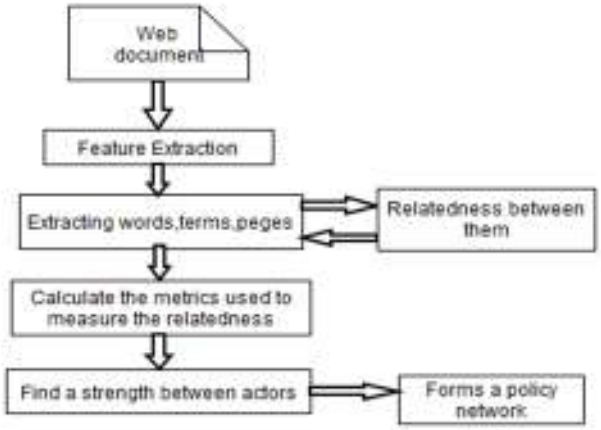

Fig.1 System Architecture

b. RELATEDNESS METRIC

i. Page-Count-Based Metrics

The set of all documents indexed by a search engine is denoted as {D}, and the cardinality of this set is denoted as |D| the set of documents that are indexed by an actor ai we use the notation $\{D_{ai}\}$. In similar fashion, the set of documents that contain two actors, a_i and a_j, is denoted as $\{D_{ai,aj}\}$ with cardinality $|D_{ai,aj}|$

Jaccard coefficient:

s_j^p Between actor's a_i and a_j is defined as follows:

$$s_j^p(a_i, a_j) = \frac{|D_{ai,aj}|}{|D_{ai}| + |D_{aj}| - |D_{ai,aj}|} \quad (1)$$

Dice coefficient :

This coefficient is closely related to the Jaccard coefficient and it is defined as

$$S_D^p(a_i, a_j) = \frac{2|D_{ai,aj}|}{|D_{ai}| + |D_{aj}|} \quad (2)$$

is equal to 1 and 0, for absolute similarity and dissimilarity, respectively.

Google based semantic relatedness :

The "normalized Google distance" is another page-count-based similarity metric that was proposed in, defined as follows:

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

This metric is a dissimilarity measure, i.e., as the distance between two actors increases the metric takes smaller values. The scores assigned by (4) are unbounded, ranging from 0 to 1. In, a variation of the normalized Google distance was used, proposing a bounded similarity measure called "Google-based semantic relatedness," defined as W - Is the context window length ,N - Is the vocabulary size, $v_{ai,j}$ - The value of can be a function of the frequency of occurrence of **vj** in the context of **ai**

3.2.2 Text-Based Metrics Computation

$$S_W^T(ai, aj) = \frac{\sum_{i=1}^{N} v_{ai,l} \cdot v_{aj,l}}{\sqrt{\sum_{i=1}^{N} (v_{ai,l})^2} \sqrt{\sum_{i=1}^{N} (v_{aj,l})^2}} \quad (4)$$

Where W is the context window length and N is the vocabulary size. The cosine similarity metric assigns 0 similarity score when **ai, aj** share on common context (completely dissimilar actors), and 1 for identical actors (orators sharing the same contexts).

3.2.3 Link-Based Metrics Computation

$$S_R^L(a_i, a_j) = \frac{\max\{\log|D_{ai}|, \log|D_{aj}|\} - \log|D_{ai,aj}|}{\log|D| - \min\{\log|D_{ai}|, \log|D_{aj}|\}}$$

Where $\{O_{ai}\}, \{O_{aj}\}$, and $\{O_{ai,aj}\}$ are the set of out links for actors **ai, aj**and jointly for both **ai** and **aj** .

3.2.4 Linear Fusion of Relatedness Metrics

$$S(a_i, a_j) = \lambda p S^P(\text{ai, aj}) + \lambda_T S^T(\text{ai,aj}) + \lambda_L S^L(\text{ai, aj})$$

$\lambda p, \lambda_L, \lambda_T$. Are the corresponding weights.Where S^P, S^T, S^L refer to the proposed page-count, text, and link-based metrics, respectively, and $\lambda p, \lambda_L, \lambda_T$ are the corresponding weights.

Two cases are investigated:

Equal weights (that sum up to 1) and inverse variance weighting (informative fusion). For informative fusion, the weights for each type of metric set equal to the inverse variance, e.g., $\lambda_P = 1/\sigma^2 p$.

The variance is computed across the relatedness scores for all actor pairs and a specific metric. **kmin, kmax**- Is the min and max scores H,E - Sample mean ,A - Is the number of actors (vertices in the network), $W_{i,j}$- is the weight (rating) of the relation (edge) between actors ai, aj.

IV. CONCLUSION

A conclusion In this work, we have shown that it is possible to automatically compute the strength of relations between actors to automatically create policy networks. A variety of feature were proposed and evaluated that used information automatically extracted from the WWW (World Wide Web). Specifically, we investigated the use of page counts, lexical context, and outlinks, as well as, their fusion, as potential features for estimating relatedness between actor pairs.

This work is a first step toward creating algorithms and tools useful to policy network analysts. Future work should invole Investigating the applicability of the proposed metrics for other types of social networks.

REFERENCES

[1] The odosis Moschopoulos, Elias Iosif, Student Member, IEEE, Leeda Demetropoulou, Alexandros Potamianos, Senior Member, IEEE, and Shrikanth (Shri) Narayanan, Fellow, IEEE, "Toward the Automatic Extraction of PolicyNetworks Using Web Links and Documents",IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 10, OCTOBER 2013.

[2] P. Kenis and V. Schneider, Policy analysis and Policy Networks: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 Assumptions and Dependencies Statistical Analysis of Political Text," Political Analysis, vol. 16,no. 4, pp. 351-355, 2008.

[5] M. Laver, K. Benoit, and J. Garry, "Extracting Policy Position 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 2002 Elections - A Research Note," Irish Political Studies, vol. 18, no. 1, pp. 97-107, 2003.

[7] J.B. Slapin 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, B. Pang, 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 L. Zhu. 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 and K. Moilanen, "Leveraging Textual Sentiment Analysis with Social Network Modelling: Sentiment Analysis ofPolitical Blogs in the 2008 U.S Presidential Election," Proc. 'FromText to Political Positions' Workshop, 2010.

[11] B. Monroe, M. Colaresi, and K. Quinn, "Fightin' Words: LexicalFeature Selection and Evaluation for Identifying the Content ofPolitical Conflict," Political Analysis, vol. 16, no. 4, pp. 372-403,2008.