Q: Please put title name here

Deokwoo Jung [#1], Deamin Park [*2]

[#] *Advanced Digital Sciences Center, Singapore*
[1] deokwoo.jung@adsc.com.sg, [2]

[*] *Korea Press Foundation, Korea Republic*
[2] heathe0@gmail.com

***Abstract*—paper abstraction is here...**
[TODO: to be done by DM]

This paper says blabla...

I. Introduction

[TODO: to be done by DM]
Introduction is here...
Q: Motivation and background of this paper ?
tanswrr ....
Q: What is the problem is why you need to solve the problem ?
Q: Any previous research works to solve the problem and why they are not sufficient if exist ?
Q: Describe your approach to solve the problem, and how can you justify the approach ?
Q: How you will demonstrate the effectiveness and the novelty of your approach ?

II. Related Work

[TODO: to be done by DM]
related work is here [1]

III. Problem Overview

[TODO: to be done by DM and DW]
Problem overview is here

IV. Model Formulation

[TODO: to be done by DW]
To formulate a model for our problem, let us first introduce a number of definitions as follows. Let a set of news article and news sources be denoted by $\mathbf{a} = \{a_1, \cdots, a_l\}$ and $\mathbf{s} = \{s_1, \cdots, s_m\}$ respectively. We assume the the total number of articles and news sources are fixed by l and m for our analysis. Individual news sources are uniquely identified by their full name, organization name and position title which defined by a 3-tuple vector $s_i = (name, organization, position)$.

Let define an association matrix between news sources S and articles A by $\mathbf{U}_{l\times m}$ whose i_{th} column and k_{th} row element $u_{ij} = 1$ if $s_i \in a_j$ and 0, otherwise.

News sources can be classified into a number of unknown categories such as economy, politics so on for example. However, the actual categories would be unknown which can be only characterized by analyzing a large volume of news articles. Let us define an undirected graph of news sources $\mathcal{G}_s$

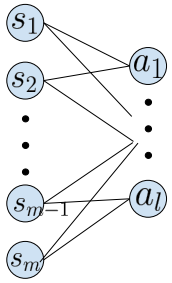

Fig. 1: An illustrated example of association graph between news sources and articles

to model a pairwise dependency of news sources where its edge represents a degree of similarity in categories shared by the pair. Empirically drawn from journalism perspective, it can be assumed that news sources are in the same category if they are shared in the same article. The empirical statement can be reduced into a structure of $\mathcal{G}_s$ whose connectivity matrix is simply computed by $\mathbf{S}_{m\times m} = \mathbf{U}\mathbf{U}^T$.

Then let us define a set of quotations in articles denoted by $\mathbf{q} = \{q_1, \cdots, a_n\}$ where n is the total number of quotations uniquely found in articles. We can quantify the similarity between two quotations by statistical occurrences of key nones. Let $\hat{q}_i$ denote a project of q_i into n-dimensional Euclidian space E_n. Then we define a distance of matrix for q_is denoted by $\mathbf{D}_q$.

Similarly to aforementioned modeling for association matrix $\mathbf{U}_{l\times m}$, let us $\mathbf{V}_{m\times n}$ and $\mathbf{Z}_{n\times l}$ denote a association matrix for news sources $-$ quotations and quotations $-$ articles, respectively. Then we can define $\mathbf{Q}_v = \mathbf{V}\mathbf{V}^T$ and $\mathbf{Q}_z = \mathbf{Z}\mathbf{Z}^T$ which are a undirected graph of quotations computed from $\mathbf{V}$ and $\mathbf{Z}$, respectively.

Then we can formulate an abstractive distance of q_is in E_n as following,

$$\mathbf{D}_q^* = w_d\mathbf{D}_q + w_v\mathbf{Q}_v + w_z\mathbf{Q}_z$$

where w_x are a weight coefficient for x.

Then finally we can classify quotations of q_is into a number of clusters given $\mathbf{D}_q^*$ using supervised clustering algorithms such as affinity propagations.

V. Algorithm

[TODO: to be done by DW]
algorithm is here

VI. Experiment

[TODO: to be done by DW]

Experiment is here

VII. Conclusion and Future Work

conclusion of this work is here.

References

[1] G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978.