International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Study on a Smoothing and Relaxation Method of Optimization Problem for Clustering of Bibliometric Networks

Myong-Hyok Jang[1], Yong Kim[2]

[1]Pyongsong Industrial University, Pyongsong, D.P.R. of Korea

[2]Department of Mathematics, University of Science, Pyongyang, D.P.R. of Korea

Abstract: ***Optimization problems for clustering of bibliometric networks are discrete optimization ones with discontinuous objective functions. Therefore, we can not apply the traditional optimization methods to solve in these problems. In this paper, we studied smoothing and relaxation method of optimization problems for clustering bibliometric networks.***

Keywords: discrete discontinuous optimization problem, clustering, smoothing method, relaxation method

1. Introduction

Bibliometrics is the scientific field that quantitatively studies all kinds of bibliographic data such as titles, keywords, authors, and cited references of papers and booklets. Bibliometric study focuses on network analysis of documents, keywords, authors and journals. The analysis of bibliometric data are mainly used in mapping and clustering.

In the several previous literatures, methods for hierarchical clustering of bibliometric networks have been studied.

A factor analysis method was studied as one of clustering methods in [2,5,8].

And clustering methods based on module functions were studied in [6,7].

An integration method of mapping and clustering was proposed in the optimization problem of the following function providing the integration method of mapping and clustering. [4].

$$\min V_1(x_1,\cdots,x_n)=\sum_{i<j}S_{ij}d_{ij}^2-\sum_{i<j}d_{ij} \qquad (1)$$

In the case of clustering, d_{ij} is

$$d_{ij}=\begin{cases}0, & x_i=x_j\\ \frac{1}{r}, & x_i\neq x_j\end{cases} \qquad (2)$$

where r is called partition parameter.

The bigger the value of this parameter is the more the number of clusters is. S_{ij} is the quantity that represents the relatedness between nodes i and j, which is given as follows:

$$S_{ij}=\frac{2mC_{ij}}{C_iC_j} \qquad (3)$$

where C_{ij} is the number of links between nodes i and j, C_i denotes the total number of links of node I, and m denotes the total number of links in the network, that is,

$$C_i=\sum_{j\neq i}C_{ij}, \qquad m=\frac{1}{2}\sum_i C_i \qquad (4)$$

In case of clustering, we need to find for each node i, a positive integer x_i that indicates the cluster to which node i belongs.

The method in [4] contains the methods proposed in [6,7] as the special case.

In the literature [4], it proved that the optimization problem of function V_1 is equivalent to the optimization problem of the function given as follows:

$$V(x_1,\cdots,x_n)=\frac{1}{2m}\sum_{i<j}\delta(x_i,x_j)w_{ij}\left(C_{ij}-r\frac{C_iC_j}{2m}\right) \qquad (5)$$

$$\delta(x_i,x_j)=\begin{cases}1, & x_i=x_j\\ 0, & x_i\neq x_j\end{cases} \qquad w_{ij}=\frac{2m}{C_iC_j}$$

The function $V(x_1,\cdots,x_n)$ is discontinuous function. Therefore, we can not use the traditional optimization methods. We discuss the following problem that makes the number of clusters less than k+1.

$$\begin{aligned}&\min V(x_1,\cdots,x_n)\\ &s.t.\ \ x\in X\end{aligned} \qquad (6)$$

where $X=\left\{x\,|\,0\le x_i\le k,\ \ x_i:\text{integer},\ \ i=\overline{1,n}\right\}$.

The paper is composed of as follows.

In Section 2 we propose a smoothing approximation of the function $V(x_1,\cdots,x_n)$ and estimate the difference between functions $V(x_1,\cdots,x_n)$ and its approximation. In Section 3 we propose a relaxation method for smoothing approximation matter of the Problem (6), and then consider one solving method of the relaxed problem. In Section 4 we consider a moving balls approximation method for this problem (11).

2. Smoothing of function $V(x_1,\cdots,x_n)$

Now let's we introduce the following smoothing approximation of the function $\delta(x_i,x_j)$:

$$f(x_i,x_j)=e^{-c(x_i-x_j)^2}$$

where c is positive parameter.

$f(x_i,x_j)$ is equal to the $\delta(x_i,x_j)$=1 when $x_i=x_j$.

In case of $x_i\neq x_j$, by making the parameter c larger we can let the $f(x_i,x_j)$ closer to the $\delta(x_i,x_j)$=0.

For simplicity, we use the notation

$$\alpha_{ij}=w_{ij}(C_{ij}-r\frac{c_ic_j}{2m}).$$

We replace $\delta(x_i,x_j)$ with $f(x_i,x_j)$ and then discuss the smooth function

$$\tilde{V}(x_1,\cdots,x_n)=\frac{1}{2m}\sum_{i<j}\alpha_{ij}f(x_i,x_j).$$

Now we consider the following problem:

$$\min \tilde{V}(x_1,\cdots,x_n)$$
$$s.t.\ \ x\in X \qquad (7)$$

Theorem 1. The difference $\left|\tilde{V}(x)-V(x)\right|$ uniformly converges to 0 as $c\to\infty$ in X.

Proof. Notice that $(x_i-x_j)^2\geq 1$ in case of $x_i\neq x_j$ since x_i,x_j are integers. Therefore, for any $x\in X$ we have

$$\left|\tilde{V}(x)-V(x)\right|\leq\frac{1}{2m}\sum_{i<j}\left|\alpha_{ij}\right|\left(f(x_i,x_j)-\delta(x_i,x_j)\right)=$$

$$=\frac{1}{2m}\sum_{\substack{i<j\\x_i\neq x_j}}\left|\alpha_{ij}\right|f(x_i,x_j)=\frac{1}{2m}\sum_{\substack{i<j\\x_i\neq x_j}}\left|\alpha_{ij}\right|e^{-c(x_i-x_j)^2}$$

$$\leq\frac{1}{2m}\sum_{\substack{i<j\\x_i\neq x_j}}\left|\alpha_{ij}\right|e^{-C}\leq\frac{1}{2m}\sum_{i<j}\left|\alpha_{ij}\right|e^{-C}\to 0\quad(C\to\infty)\ \blacksquare$$

Let x^*, $\tilde{x}(c)$ be the solutions of problems (6) and (7), respectively.

The following theorem shows the upper bound of difference between $\tilde{V}(\tilde{x}(c))$ and $V(X^*)$, and this difference converges to 0 under $c\to\infty$.

Theorem 2. Let $\alpha_{ij}\geq 0$ for any i and j.Then we have

$$\tilde{V}(\tilde{x}(c))\to V(x^*)\qquad(c\to\infty)\qquad(8)$$

$$V(\tilde{x}(c))\to V(x^*)\qquad(c\to\infty)\qquad(9)$$

$$\tilde{V}(\tilde{x}(c))-V(x^*)\leq\frac{1}{2m}\sum_{\substack{i<j\\x_i\neq x_j}}\left|\alpha_{ij}\right|e^{-C}\qquad(10)$$

Proof. From the definition of $\tilde{V}(x)$, if $\alpha_{ij}\geq 0$ then $V(x)\leq\tilde{V}(x)$ for any i,j.

Hence we have

$$V(x^*)\leq V(\tilde{x}(c)\leq\tilde{V}(\tilde{x}(c)\leq\tilde{V}(x^*)\qquad(11)$$

On the other hand, from the theorem 1, $\tilde{V}(x^*)-V(x^*)\to 0\quad(c\to\infty)$.

Hence, $\tilde{V}(\tilde{x}(c))\to V(x^*)$, $V(\tilde{x}(c))\to V(x^*)\ (c\to\infty)$, that is, (8) and (9) hold.

From(11) the inequality $\tilde{V}(\tilde{x}(c))-V(x^*)\leq\tilde{V}(x^*)-V(x^*)$ hold, and therefore from the theorem 1 we have

$$\tilde{V}(\tilde{x}(c))-V(x^*)\leq\tilde{V}(x^*)-V(x^*)\leq\frac{1}{2m}\sum_{\substack{i<j\\x_i\neq x_j}}\left|\alpha_{ij}\right|e^{-C}$$

Hence the inequality (10) hold.

Let us discuss the general case where condition $\alpha_{ij}\geq 0,\ \ \forall i,j$ is not satisfied.

Lemma 1 ([3]) Let $X\subset R^n$ be a closed set, $\Phi,\Psi:R^n\to R$ be continuous functions, and assume the following inequality holds:

$$\left|\Phi(x)-\Psi(x)\right|\leq\delta,\quad\forall x\in X.$$

If $x^*,\tilde{x}$ are the gloval minimizers of Φ, Ψ in X, respectively, the following inequalities hold for:

1) $\left|\Phi(x^*)-\Psi(\tilde{x})\right|\leq\sup_{x\in X}\left|\Phi(x)-\Psi(x)\right|\leq\delta$

2) $\left|\Phi(x^*)-\Phi(\tilde{x})\right|\leq 2\delta,\quad\left|\Psi(x^*)-\Psi(\tilde{x})\right|\leq 2\delta$.

Using this lemma, we can get the following result.

Theorem 3. The following inequality holds for:

$$\left|\tilde{V}(\tilde{x}(c))-V(x^*)\right|\leq\frac{1}{2m}\sum_{i<j}\left|\alpha_{ij}\right|e^{-C}.$$

Proof. Let us denote $\delta=\frac{1}{2m}\sum_{i<j}\left|\alpha_{ij}\right|e^{-C}$ then from theorem 1 we have

$$\left|\tilde{V}(x)-V(x)\right|\leq\delta,\quad\forall x\in X.$$

If, therefore, we apply lemma, the result of lemma 1 is obtained.

Our above results show that if c is large, then the optimal value of smooth integer programming problem (7) can be enough close to the optimal value of original problem (6).

3. Relaxation of problem (7)

Considering the formula of w_{ij}, the objective function $\tilde{V}(x_1,\cdots,x_n)$ can be rewrited as follows:

$$\tilde{V}(x)=\frac{1}{2m}\sum_{i<j}\beta_{ij}e^{-c(x_i-x_j)^2}$$

where $\beta_{ij}=(r-\frac{2mc_{ij}}{c_i c_j})$ (it is clear that $\alpha_{ij}=\beta_{ij}$).

We assume $\beta_{ij}\geq 0,\quad \forall i<j$.

Let us $\tilde{X}\overset{D}{=}\{x\in R^n \mid \ 0\leq x_i\leq k,\quad i=\overline{1,n}\}$ and then consider the following relaxation problem for (7):

$$\min \tilde{V}(x)$$
$$s.t.\, x\in\tilde{X} \qquad (11)$$

First, we estimate the difference between the minimum values of $\tilde{V}(x)$ in $\tilde{X}$ and X, respectively.

Theorem 4. Let $\tilde{x}$ be the minimizer of $\tilde{V}(x)$ in $\tilde{X}$ and x^* is the minimizer of $\tilde{V}(x)$ in X. Then we have

$$\tilde{V}(x^*)-\tilde{V}(\tilde{x})\leq\frac{1}{2m}\sum_{i<j}\beta_{ij}\left(e^{-c(|\tilde{x}_i-\tilde{x}_j|-1)^2}-e^{-c(\tilde{x}_i-\tilde{x}_j)^2}\right) \qquad (12)$$

Proof. Assuming each component of $\tilde{x}=(\tilde{x}_1,\cdots,\tilde{x}_n)$ is denoted as

$$\tilde{x}_i=[\tilde{x}_i]+\rho_i,\quad i=\overline{1,n},$$

where [] denotes the integer-part. We then define $\tilde{y}=(\tilde{y}_1,\cdots,\tilde{y}_n)$ as follows:

$$\tilde{y}_i\overset{D}{=}\begin{cases}[\tilde{x}_i]+1, & \rho_i>0.5\\ [\tilde{x}_i], & \rho_i\leq 0.5\end{cases}\qquad i=\overline{1,n}$$

It is clear

$$|\tilde{y}_i-\tilde{x}_i|\leq 0.5,\quad i=\overline{1,n}.$$

Now, inequalities: $\tilde{V}(x^*)\leq\tilde{V}(\tilde{x})$ hold, for, hence

$$\tilde{V}(x^*)-\tilde{V}(\tilde{x})\leq\tilde{V}(\tilde{y})-\tilde{V}(\tilde{x})=$$
$$=\frac{1}{2m}\sum_{i<j}\beta_{ij}e^{-c(\tilde{y}_i-\tilde{y}_j)^2}-\frac{1}{2m}\sum_{i<j}\beta_{ij}e^{-c(\tilde{x}_i-\tilde{x}_j)^2}$$
$$=\frac{1}{2m}\sum_{i<j}\beta_{ij}\left(e^{-c(\tilde{y}_i-\tilde{y}_j)^2}-e^{-c(\tilde{x}_i-\tilde{x}_j)^2}\right)$$
$$=\frac{1}{2m}\sum_{i<j}\beta_{ij}\left(e^{-c|\tilde{y}_i-\tilde{y}_j|^2}-e^{-c(\tilde{x}_i-\tilde{x}_j)^2}\right)$$
$$\leq\frac{1}{2m}\sum_{i<j}\beta_{ij}\max\begin{Bmatrix}e^{-c(|\tilde{x}_i-\tilde{x}_j|-1)^2}-e^{-c(\tilde{x}_i-\tilde{x}_j)^2},\\ e^{-c(|\tilde{x}_i-\tilde{x}_j|-1)^2}-e^{-c(\tilde{x}_i-\tilde{x}_j)^2}\end{Bmatrix}$$
$$=\frac{1}{2m}\sum_{i<j}\beta_{ij}\left(e^{-c(|\tilde{x}_i-\tilde{x}_j|-1)^2}-e^{-c(\tilde{x}_i-\tilde{x}_j)^2}\right).$$

4. A Moving Balls Approximation Method for the Problem (11)

In this section we consider a moving balls approximation method for the problem (11). In the paper [1] authors introduced the moving balls approximation method(MBA) for a class of smooth constrained minimization problems. In this paper we adapt this method for the problem (11) and analyse the convergence of it.

The Laglangian function of the problem (11) is

$$L(x,\lambda,M)=\tilde{V}(x)+\lambda^T(x-k)-M^T x,$$
$$\lambda\geq 0,\qquad M\geq 0$$

where $\lambda=(\lambda_1,\cdots,\lambda_n),\ M=(M_1,\cdots,M_n)$.

Let S denote set of KKT points of the problem (11), that is,

$$S=\{x\in\tilde{X}:\exists\lambda_i\geq 0, M_i\geq 0,\ \nabla\tilde{V}(x)+\lambda-M=0,$$
$$\lambda_i(x_i-k)=0, M_i x_i=0, i=1,\cdots,n\}$$

Lemma 2. ([1]) Let $g:R^n\to R$ be a continuously differentiable function with global Lipschitz gradient and Lipschitz constant L_g. Then for any $L'\geq L_g$,

$$g(y)\leq g(x)+(y-x)^T\nabla g(x)+\frac{L'}{2}\|x-y\|^2,\quad \forall x,y\in R^n$$

In [1] authors considered the problems

$$f_*=\inf\{f(x):x\in F\},$$
$$F=\{x\in R^n: f_i(x)\leq 0, i=1,\cdots,m\}$$

and made assumptions that the functions $f(x), f_i(x), i=1,\cdots,m$ are continuously differentiable with Lipschitz gradients.

We start from estimating of Lipschitz constant of $\nabla\tilde{V}(x)$.

First, let $\varphi_{ij}(x)=e^{-c(x_i-x_j)^2}$ and estimate Lipschitz constant of $\nabla\varphi_{ij}(x)$.

Since
$$\nabla\varphi_{ij}(x)=(0,\cdots,-2e^{-c(x_i-x_j)^2}2(x_i-x_j),0,\cdots,$$
$$0,2e^{-c(x_i-x_j)^2}(x_i-x_j),0,\cdots,0)$$

, so we have

$$\|\nabla\varphi_{ij}(x)-\nabla\varphi_{ij}(y)\|\le$$
$$\le\left|-2e^{-c(x_i-x_j)^2}(x_i-x_j)+2e^{-c(x_i-x_j)^2}(y_i-y_j)\right|+$$
$$+\left|2e^{-c(x_i-x_j)^2}(x_i-x_j)-2e^{-c(x_i-x_j)^2}(y_i-y_j)\right|\le$$
$$\le\left|-2e^{-c(x_i-x_j)^2}(x_i-x_j)\right|+\left|2e^{-c(x_i-x_j)^2}(y_i-y_j)\right|+$$
$$+\left|2e^{-c(x_i-x_j)^2}(x_i-x_j)\right|+\left|2e^{-c(x_i-x_j)^2}(y_i-y_j)\right|\le$$
$$\le 4|x_i-x_j|+4|y_i-y_j|\le 8\|x-y\|$$

Define $\beta=\max_{i<j}\{\beta_{ij}\}$ and $Lv=\frac{8}{2m}\frac{n(n-1)}{2}\beta$, then L_V is Lipschitz constant of $\nabla\tilde{V}(x)$.

Now let $x\in\tilde{X}$ is given, and then we define the function $h(x,y)$ as follows:

$$h(x,y)=\tilde{V}(x)+\nabla\tilde{V}(x)^T(y-x)+\frac{L_V}{2}\|x-y\|^2,$$

where L_V is above Lipschitz constant of $\nabla\tilde{V}(x)$.

This is strongly convex function of y for each fixed x. Therefore the problem

$$\min\ h(x,y) \qquad s.t.\ \ y\in\tilde{X} \qquad (13)$$

has the unique solution. Let $p(x)=(p_1(x),\cdots,p_n(x))$ denote this unique solution.

Lemma 3. Let $x\in\tilde{X}$. Then $x\in S$ if and only if $x=p(x)$.

Proof. Since linear function constraint qualification (LFCQ) holds for the problem (13), $p(x)$ satisfies the following KKT condition:

$$\exists\quad u=(u_1,\cdots,u_n)\in R_+^n,$$
$$v=(v_1,\cdots,v_n)\in R_+^n$$
$$\nabla\tilde{V}(x)+L_V(p(x)-x)+u-v=0 \qquad (14)$$
$$u_i(p_i(x)-k)=0 \qquad i=\overline{1,n} \qquad (15)$$
$$v_i p_i(x)=0 \qquad i=\overline{1,n} \qquad (16)$$

If $p(x)=x$, then (14)~(16) are KKT conditions for the problem [11].

Hence, $x\in S$.

Now let $x\in S$. Then the point $y=x$ satisfies the KKT condition for the problem (13). On the other hand, the problem (13) is convex programming problem, so the KKT condition is sufficient condition for optimality. Hence the point x is a solution of the problem (13). Since $p(x)$ is the unique solution of the problem (13), therefore $y=x=P(x)$.

Lemma 4. The mapping $p(x)$ is continuous at $\tilde{X}$.

The proof can be done likewise in [1].

Algorithm

Step 1: Choose a point $x^0\in\tilde{X}$. Set k=0.

Step 2: For $k=1,2,\cdots$, repeat the following procedure:

$$x^k:=p(x^{k-1})$$

Let's discuss the convergence of this algorithm.

Theorem 5. Let $\{x^k\}$ be the sequence generated by the algorithm. Then

① the sequence of function values $\{\tilde{V}(x^k)\}$ is monotonically decreasing,

② any limit point of the sequence $\{x^k\}$ is in S.

Proof. ① If $x^{k-1}=p(x^{k-1})$ for some $k\ge 1$, then $x^{k-1}\in S$ and thus the theorem is proved. Let's assume that $x^{k-1}\ne p(x^{k-1})$ for all $k\ge 1$. From lemma 2,

$$\tilde{V}(x^k)\le\tilde{V}(x^{k-1})+\nabla\tilde{V}(x^{k-1})^T(x^k-x^{k-1})+\frac{L_V}{2}\|x^k-x^{k-1}\|^2 \quad (17)$$

Since $p(x)$ is the solution of the problem (13), from the optimality condition for this problem we have

$$(y-p(x))^T\left[\nabla\tilde{V}(x)+L_V(p(x)-x)\right]\ge 0,\quad \forall y\in\tilde{X}.$$

Hence, considering $x^k=p(x^{k-1})$, we have

$$\nabla\tilde{V}(x^{k-1})^T(x^k-x^{k-1})\le -L_V\|x^k-x^{k-1}\|^2 \qquad (18)$$

From (17) and (18),it follows that

$$\tilde{V}(x^k)\le\tilde{V}(x^{k-1})-\frac{L_V}{2}\|x^k-x^{k-1}\|^2 \qquad (19)$$

Therefore, since $x^k=p(x^{k-1})\ne x^{k-1}$, the sequence $\{\tilde{V}(x^k)\}$ is monotonically decreasing.

② It is clear that $\tilde{X}$ is compact set and therefore $\{x^k\}\subset\tilde{X}$ is bounded. Hence, all limit points of $\{x^k\}$ is in $\tilde{X}$.

Passing to the limit in (19), we get

$$\lim_{k\to\infty}\|p(x^{k-1})-x^{k-1}\|=0.$$

Hence, by the continuity of $P(x)$, the equality $P(x^*)=x^*$ is true for any limit point x^* of $\{x^k\}$. Therefore $x^*\in S$.

References

[1] Auslender A., Shefi R., Teboulle M., A moving balls approximation method for a class of smooth constrained minimization problems, SIAM J. Option, V. 20, No. 6, 3232~3259, 2010.

[2] Boyack K. W., Klavans R., Borner K., Mapping the backbone of science, Scientometrics, 64(3), 351~374, 2005.

[3] Jorg F., Huifu X., Stochastic multiobjective optimization: sample average approximation and applications, J. Optim theory Appl., 151, 135~162, 2011.

[4] Nees Jan Van Eck, Methodological advances in biblimetric mapping of science, 2011.
[5] Leydesdorff L., Rafols I., A global map of science based on the ISI subject categories, Journal of the American society for information science and technology, 60(2), 348~362, 2009.
[6] Wallace M. L., Gingras Y., Duhon R., A new approach for detecting scientific specialties from raw cocitation networks, Journal of the American Society for information science and Technology, 60(2), 240~246, 2009.
[7] Zhang L., Lin X.,Liang L.,Glanzel W., Subject clustering analysis based on ISI category classification, Journal of informetrics, 4(2), 185~193, 2010.
[8] Zhao D., Strotmann A., Information science during the first decade of the web: An enriched author cocitation analysis, Journal of the American Society for information science and technology, 59(6), 916~937, 2008.