**MP-lasso chart: Multi-level polar chart for visualizing group lasso analysis of genomic data (가칭)**

**Min Song1, Taesung Park2, Mira Park3\***

1Department of Statistics, Korea University, Seoul, 02841, Republic of Korea

2Department of Statistics, Seoul National University, Seoul, 08826, Republic of Korea

3Department of Preventive Medicine, Eulji University, Daejeon, 34824, Republic of Korea

**\***Correspondence author

Email addresses:

MS: sm0684@korea.ac.kr

TP: tspark@stats.snu.ac.kr

MP: mira@eulji.ac.kr

**Abstract (박미라가 쓸것임)**

~~To identify pathways associated with survival phenotypes using gene expression data, we recently proposed the hierarchical structural component model for pathway analysis of gene expression data (HisCoM-PAGE) method. The HisCoM-PAGE software can consider hierarchical structural relationships between genes and pathways and analyze multiple pathways simultaneously. It can be applied to various types of gene expression data, such as microarray data or RNA sequencing data. We expect that the HisCoM-PAGE software will make our method more easily accessible to researchers who want to perform pathway analysis for survival times.~~

**Keywords:**

Group lasso analysis, Sparse group lasso analysis, Multi-level Polar Chart, Variable Selection, Group structure visualization

**Availability:** Multi-level Polar Chart software is available on the website (<http://github> 에 넣자)

**Introduction (박미라가 쓸예정)**

Gwas 분석에서 Lasso analysis의 유용성에 대해서 언급

Lasso의 종류: 그룹라소, sparse lasso….(레퍼런스들 필요)

시각화의 의미: 유용성, 기존의 것과 차별성 등.

R로 만들었다.

**Implementation (송민이 쓸 예정: 다른 논문 용어 참고해서 영어로 쓰면 좋고 어려우면 한글로라도 정리)**

The workflow of the The MP-lasso chart software is shown in Fig. 1. The MP-lasso chart performs visualizing variable selection in group structure penalty function. This MP chart need simply cross-validated object from group structure penalty method(group lasso or sparse group lasso) and then return pie shaping interactive plot. The MP-lasso chart software is entirely written in R code.

Input

The MP-lasso chart software takes the following two inputs: (1) a cross-validated object, in which output object of group structure penalty method from R function, contains regression coefficients according by sequencial lambda values and optimal lambda value. Each R function needed shown in Table 1. (2) a group name, where represents group structure for each variables, requires character type or integer type used in fitting cross-validated object. Users do not need to prepare data matrix again, handle several options, find regression coefficients for optimal lambda value.

그냥 간단하게 두 개의 input만 필요하다고 말하고 싶음. 왜냐하면, group lasso와 sparse group lasso 함수인 gglasso(), sgl()의 input이 대강 똑같지 완벽히 같은 건 아니어서 각각 help찾으면서 fitting을 해야함. 예를 들어 y에 대한 코딩도 -1 or 1이냐 0 or 1이냐 달라서 data matrix도 다르게 만들어줘야함. 이러한 면에서 MP-chart는 유저 입장에서 사용하기 쉽게 만들어졌다. 다른 방법론을 그림으로 보고 싶다고해서 input, argument가 바뀌는 것이 아니니까.

Execution

Plotting procedure consist of two stage

For the first level of stage, graph groups according importance. Suppose k group is selected by group selection method. Draw a circle and divide the angle equally by k and calculate absolute average of coefficients in each group(or absolute maximum of coefficients in each group). after that sort this value desceding order and mapping to radius of circle in each group. then for second level of stage, mapping absolute value of coefficients to each fan-shaped circle interactively with points which contain name of coefficients and orginal scale of coefficients.

The detailed function and arguments as follow and also shown in Table 1.

\* PieBW\_gglasso(cv\_object=NULL, group = NULL, lambda.type = "min", sort.type = "mean")

- Description : Drawing MP-lasso from a fitted cv.gglasso() object

- Arguments :

cv\_object : fitted cv.gglasso() object.

group : group name, character or integer.

lambda.type : selection of optimal lambda value. if lambda.type=”min”, then select

lambda vaule which gives minimum loss for model. if lambda.type=”1se”, then select

lambda vaule which gives minimum loss in one stadard error range.

sort.type : selection of sort type for group importance. if sort.type=”min”, then group sorted by absolute average of coefficients in each group. if sort.type=”max”, then group sorted by absolute average of coefficients in each group.

\* PieBW\_gglasso(cv\_object=NULL, group = NULL, lambda.type = "min", sort.type = "mean")

- Description : Drawing MP-lasso from a fitted cvSGL() object

- Arguments :

cv\_object : fitted cvSGL() object.

group : group name, character or integer.

lambda.type : selection of optimal lambda value. if lambda.type=”min”, then select

lambda vaule which gives minimum loss for model. if lambda.type=”1se”, then select

lambda vaule which gives minimum loss in one stadard error range.

sort.type : selection of sort type for group importance. if sort.type=”min”, then group sorted by absolute average of coefficients in each group. if sort.type=”max”, then group sorted by absolute maximum of coefficients in each group.

Output

Interpretations

**Table 1 coverage of MP-Lasso chart**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | R function | CV object | R Package for CV | method reference |
| Group lasso | PieBW\_gglasso() | cv.gglasso() | gglasso |  |
| Sparse group lasso | PieBW\_SGL() | cvSGL() | SGL |  |
| **기타옵션들 채워보기** |  |  |  |  |
|  | Network?? Lasso, |  |  |  |

**워크플로우그리기**

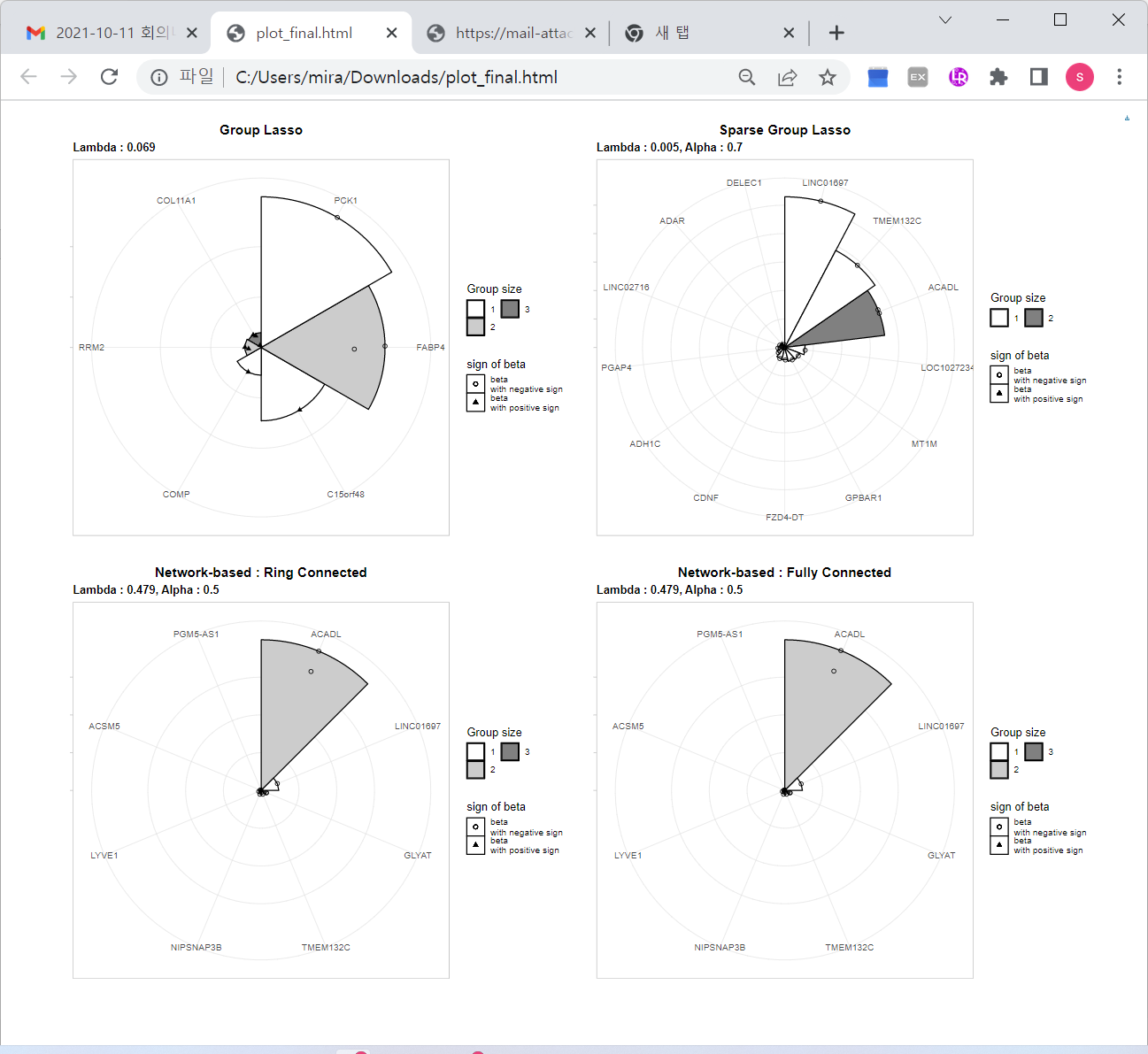
Fig. 1. The workflow of multi-level polar chart for group lasso analysis of genomic data. ~~The pathway is constructed by a weighted sum of its mapped genes. The pathway coefficient is represented as β and the gene coefficient is represented as w in the Figure.~~

**Results (송민이 쓸 예정: 다른 논문 용어 참고해서 영어로 쓰면 좋고 어려우면 한글로라도 정리)**

To draw a MP-lasso chart, generate simulated data where n=100 and p=200. regression coefficients set as follows to divide group into higher absolute group mean(group1) and lower absolute group mean but higher coeffient

**예제데이터설명넣기 : 오픈자료이면 설명을 제대로 넣고 아니면 변수수나 샘플수등만 간단히**

**아래와 같은 식으로 결과그림넣기**



**Conclusion (박미라가 쓸것임)**

~~In this paper, we introduce the HisCoM-PAGE software for pathway analysis of the survival phenotype using gene expression data. The HisCoM-PAGE software may be a useful tool for the identification of pathways associated with the survival phenotype. The software is freely available on the website, along with a detailed tutorial.~~

**Orcid**

Min Song : 오키드 만들어서 넣기

Taesung Park:  https://orcid.org/0000-0002-8294-590X.

Mira Park: https://orcid.org/0000-0003-3827-9089

**Author contribution**

Conceptualization: MP. Data curation: MS. Funding acquisition: MP. Methodology: MP, TP. Writing - original draft: MS. Writing - review & editing: MP, TP.

**Conflicts of interest**

No potential conflict of interest relevant to this article was reported.

**Acknowledgments**

This research was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2021R1A2C1007788).

**References**

참고한 레퍼런스 다 보내주세요.