Application of Biclustering Algorithm in Bioinformatics and High-Quality Leather Products

Conference: 13th AICLST, Kolkata, 6–8 March 2026

Paper reference: TF-85/O

Printed pages: 303–308

Source: *Leather 2050: Challenges & Opportunities*, ILTA (2026)

Application of Biclustering Algorithm in Bioinformatics and

High-Quality Leather Products

Brijesh Kumar Sriwastava1, Falguni Sinhababu2, Srirupa Dasgupta3, Kalyan Mahata4,

1,2Department of Computer Science and Engineering, 3,4Department of Computer Science and Engineering,

5Department of Computer Science and Engineering,

Ref. No. TF-85/O

Ipsita Sengupta5

Government College of Engineering and Leather Technology, Kolkata-700106, India.

Abstract

Motivation: A fast-biclustering algorithm for binary datasets is presented here. In order to significantly reduce the computational overhead in comparison to the state-of-the-art BiBit algorithm that uses repeated bit-wise AND operations followed by comparisons, we have introduced a novel encoding and search strategy in our method.

We can use biclustering in the leather industry for production processes optimization, productivity enhancement, and improve supply chain management. This can be done by grouping similar inputs (like raw hides, chemicals) and conditions (for example process times, temperatures). Inspite of wide use of biclustering in bioinformatics, its application in the leather industry includes adapting the technique to manufacturing and supply chain data. One of its applications is in the leather industry is optimizing the tanning process. In order to achieve desired leather properties in the leather manufacturing technique, it uses different types of chemicals and specific conditions like temperature, duration and pH levels. This can be done if we keep different batches of raw hides in rows and several combinations of chemical agents and process parameter in columns. This data matrix can be used to find biclusters which are subset of hides and subsets of process or chemical conditions.

Correspondence author email : bsriwastava@gmail.com

In this way finding biclusters will help to find high quality soft and durable leather which in turns also minimize waste.

1. Introduction

The genes in small groups can contribute in multiple cellular processes or pathways of interest that are only active under a subset of the conditions. Biclustering has been applied in gene expression data analysis. Madeira et al. (Madeira and Oliveira, 2004) have analyzed a large number of existing approaches of biclustering, and classified them with varying types of biclusters. The classification is based on the patterns of biclusters that are revealed, the methods used to in the searching. Liu et al. have worked on OP-Cluster which looked for the Order Preserving clusters, in which each subset of genes induce a similar linear ordering along a subset of conditions(Liu, et al., 2004). FABIA have designed over a multiplicative model and Bayesian techniques, where linear dependencies between gene expression and conditions have been considered and it captures heavy tailed distributions as perceived in real-world transcriptomic data (Hochreiter, et al., 2010). DeBi algorithm is based on frequent itemset, a data mining approach. It determines homogeneous biclusters in which each gene is related with a subset of samples(Serin and Vingron, 2011). A closed itemset enumeration algorithm, BiModule which tries to find such biclusters, it is possible to select only biclusters satisfying certain criteria such as a user-specified bicluster size, an

enrichment of functional annotation terms after an exhaustive enumeration(Okada, et al., 2007). In the work of (Rodriguez-Baena, et al., 2011), a novel biclustering algorithm (BiBit) was presented for extraction of bit-patterns from binary datasets. Both synthetic and experimental datasets were used to show the efficacy of the BiBit method over the-then popular method, especially Bimax (Preliæ, et al., 2006). In this work, we developed biclustering tool which is faster than BiBit, while extracting almost all the original biclusters. In order

to significantly reduce the computational overhead in comparison to the BiBit algorithm that uses repeated bit-wise AND operations and comparisons, we have introduced a novel encoding and search strategy in this work.

Here, we have worked with decimal numbers representing four-bit binary numbers. Template patterns are first obtained by logical-AND operations between an initial seed-pair of rows. To make a faster search, we have modified the encoding strategy by introducing 16 new sets of numbers, representing all possible combination of bitwise AND operations between two four-bit binary numbers and representing their group. The developed methodology requires reduced number of computational steps in comparison to the state-of-the-art BiBit method and relevant maximal biclusters are extracted in significantly less amount of time.

Most of algorithms are capable to discover biclusters having 1’s or 0’s only but not the both of 1- biclusters and 0-biclusters simultaneously. The method proposed by (Lee and Huang, 2014) finds non-pure biclusters, and can simultaneously identify the 1-prevalent blocks and 0-prevalent blocks.

BicPAM has performed exhaustive searches to discover biclusters with some coherencies under relaxed conditions (arbitrary number and structure of biclusters). BicPAM has integrated existing some contributions on pattern-based biclustering with a methods to deal with more flexible

expression profiles and to handled varying levels of missing values and noise (Henriques and Madeira, 2014).

The algorithm HiBi uses the set of exact biclusters as the input which are results of exact biclustering algorithm. It gives the set of inexact biclusters (Michalak, et al., 2014).

Algorithm 1 To find Bit-patterns biclustering algorithm

Once a template is formed, we start to select those remaining rows in such a way that 1st column decimal value of template belongs to of 1st column decimal value of row and if match found then we select all those rows which are satisfying this criteria. In this way remaining row which does not satisfies this criteria, we left these row. Then we start with these filtered rows for 2nd column value matching and this helps to filtering remaining rows and the process is continued up to last column as last step. In this process we always take only those rows as input rows which satisfied the criteria.

Now for 1st component (column) of , let N number of comparison are required to select those rows whose 1st

column value matches as per above mentioned criteria and only number of rows will be selected for 2nd

column value matching. So, in first step of this part only number of comparisons are required. In second step i.e., for 2nd column value number of comparisons are required to check 2nd column value and number of rows are only selected. This procedure is continued and at last comparison is required.

3 Conclusions

BiBit proposed a faster biclustering approach in binary datasets with respect to the well-known Bimax algorithm (Preliæ, et al., 2006). In our work, we have developed an even faster biclustering method by introducing the novel encoding and searching steps, thereby improving the computation time significantly. We can use biclustering in the leather industry for production processes optimization, productivity enhancement, and improve supply chain management. This can be done by grouping similar inputs (like raw hides, chemicals) and conditions (for example process times, temperatures). IN spite of wide use of biclustering in bioinformatics, its application in the leather industry includes adapting the technique to manufacturing and supply chain data. One of its applications is in the leather industry is optimizing the tanning process. In order to achieve desired leather properties in the leather manufacturing technique, it uses different types of chemicals and specific conditions like temperature, duration and pH levels. This can be done if we keep different batches of raw hides in rows and several combinations of chemical agents and process parameter in columns. This data matrix can be used to find biclusters which are subset of hides and subsets of process or chemical conditions. In this way finding biclusters will help to find high quality soft and durable leather which in turns also minimize waste.

References

a) Henriques, R. and Madeira, S.C. (2014) BicPAM: Pattern-based biclustering for biomedical data analysis, Algorithms for Molecular Biology, 9, 27.

b) Hochreiter, S., et al. (2010) FABIA: factor analysis for bicluster acquisition, BIOINFORMATICS, 26, 1520-1527.

c) Lee, S. and Huang, J.Z. (2014) A biclustering algorithm for binary matrices based on penalized Bernoulli likelihood, Statistics and Computing, 24, 429-441.

d) Liu, J., Yang, J. and Wang, W. (2004) Biclustering in gene expression data by tendency, Computational Systems Bioinformatics Conference, 2004. CSB 2004. Proceedings. 2004 IEEE, 182-193.

e) Madeira, S.C. and Oliveira, A.L. (2004) Biclustering algorithms for biological data analysis: a survey, Computational Biology and Bioinformatics, IEEE/ACM Transactions on, 1, 24-45.

f) Michalak, M., Lachor, M. and Polański, A. (2014) HiBi–The Algorithm of Biclustering the Discrete Data. Artificial Intelligence and Soft Computing. Springer, pp. 760-771.

g) Okada, Y., Fujibuchi, W. and Horton, P. (2007) A biclustering method for gene expression module discovery using a closed itemset enumeration algorithm, IPSJ Digital Courier, 3, 183-192.

h) Pomeroy, S.L., et al. (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression, Nature, 415, 436-442.

i) Prelić, A., et al. (2006) A systematic comparison and evaluation of biclustering methods for gene expression data, BIOINFORMATICS, 22, 1122-1129.

j) Rodriguez-Baena, D.S., Perez-Pulido, A.J. and Aguilar, J.S. (2011) A biclustering algorithm for extracting bit-patterns from binary datasets, BIOINFORMATICS, 27, 2738-2745.

k) Salwinski, L., et al. (2004) The Database of Interacting Proteins: 2004 update, Nucleic Acids Research, 32, D449-D451.

l) Serin, A. and Vingron, M. (2011) DeBi: Discovering Differentially Expressed Biclusters using a Frequent Itemset Approach, Algorithms for Molecular Biology, 6, 18.