# Semiring Rank Matrix Factorization

Absract:

Rank data, in which each row is a complete or partial ranking of available items (columns), is ubiquitous. It has been used to represent, for instance, preferences of users, the levels of gene expression, and the outcomes of sports events. While rank data has been analysed in data mining, pattern mining in such data has so far not received much attention. To alleviate this state of affairs, we introduce a novel pattern set mining approach. It is based on a semiring matrix factorisation framework. Rather than using the traditional linear algebra for matrix factorisation we employ the max-product semiring. Our semiring matrix factorisation framework is then applied to two tasks: sparse rank matrix factorisation and rank matrix tiling. Algorithms for these tasks are proposed, and their effectiveness is demonstrated on extensive experiments using both synthetic and real data.

**Existing System:**

We develop a generic framework for unsupervised discovery of regularities (patterns) in rank data. In this type of data, each row (transaction) is a complete or a partial ranking of the available columns (items). Rank data naturally occurs in many situations of interest. Consider, for instance, cycling competitions where the items could be the cyclists and each transaction would correspond to a race, or in a business context, where the items could be particular companies and the transaction could specify the rank of their quotation for a particular service.

In general, ranking forms a natural abstraction for purely numeric data, which often arises in practice and may be noisy or imprecise. Especially when the rows are incomparable, e.g., when they contain measurements on different scales, transforming the data to rankings may result in a more informative representation.

**Proposed System:**

We proposed to use a max-product semiring defined on permissible rank values of the data to calculate the matrix product of the two factorised matrices. To mine a specific type of data regularity, we proposed to use the two factorised matrices to define the patterns of interest by constraining the values of these matrices as well as an appropriate scoring function to measure the quality of the factorisation. We demonstrated how the proposed framework can be applied on two existing rank data mining problems, namely rank matrix tiling and Sparse RMF. Modelling the two problems using this framework illustrates the expressiveness and flexibility of the approach. Experiments on both synthetic datasets and real world problems shows that the framework is capable of discovering different types of structure as well as obtaining high quality solutions.

**Modules**:

- Semiring rank matrix factorization (srmf).
- Max-product semiring rank matrix tiling.

**SYSTEM REQUIREMENTS**

**H/W System Configuration:-**

| | | |
|---|---|---|
| Processor | - | Pentium –III |
| RAM | - | 256 MB (min) |
| Hard Disk | - | 20 GB |
| Key Board | - | Standard Windows Keyboard |
| Mouse | - | Two or Three Button Mouse |
| Monitor | - | SVGA |

**S/W System Configuration:-**

| | | |
|---|---|---|
| Operating System | : | Windows95/98/2000/XP |
| Application Server | : | Tomcat5.0/6.X |

| | | |
|---|---|---|
| Front End | : | HTML, Jsp |
| Scripts | : | JavaScript. |
| Server side Script | : | Java Server Pages. |
| Database | : | MySQL 5.0 |
| Database Connectivity | : | JDBC |