**The Borda Social Choice Voting Rule**

- **Borda** is a voting rule which considers each dimension in a multi-dimensional scenario in an equal manner.
- We use Borda in k-means++ for the allocation of objects to a cluster.
- This allows more influence of smaller domains, because every candidate receives equal weighted votes from each voter.

**Definition (Borda Winner):**

Given $k$ candidates $C_i$ and $d$ voters $V_j$. A voter has to assign a vote $v_{ij} \in \{0, \ldots, k-1\}$, for $i = 1, \ldots, k$, to each candidate. All $v_{ij}$ are pairwise distinct.

1) Sum up the votes for each candidate:

$$bordaSum_{C_i} = \sum_{j=1}^{d} v_{ij}$$

2) Determine the Borda winner:

$$bordaWinner = \max\{bordaSum_{C_i} \mid i = 1, \ldots, k\}$$

**In a clustering scenario:**

- The candidates are the available clusters.
- The voters are the dimensions of the $d$-dimensional object which should be allocated to a cluster.
- Votes are assigned to the distances between the object and the centroids of the clusters.
- Closest distance gets a maximum vote of $k-1$, the second closest $k-2$, ..., the largest distance gets a vote of 0.
- Allocate the object to the Borda winner.

**Demo Architecture:**

- Web application-based recommender system using user preferences.
- Uses the Internet Movie Database (IMDb).
- Evaluation mode for user study.

**Example:**

- Bob favors old-school movies of the late 70s to the early 90s, with a runtime between 90 and 130 minutes and a user-rating higher than 7.

**Objective:**

- Allocate movie 'Die Hard 2' (27) to one of $k=3$ clusters.
- Use movies (1), (7) and (23) as initial centroids for clusters $C_1$, $C_2$ and $C_3$.

**Determine the Borda Winner:**

- 'Die Hard 2' is allocated to the Borda Winner $C_1$.
- Using the squared Euclidean distance, 'Die Hard 2' would be allocated to $C_3$.

**Table:**

<table>
<thead>
<tr>
<th>ID</th>
<th>movie</th>
<th>rating</th>
<th>time</th>
<th>year</th>
<th>genres</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Star Wars</td>
<td>8.8</td>
<td>125</td>
<td>1997</td>
<td>Action, Sci-Fi</td>
</tr>
<tr>
<td>7</td>
<td>Reservoir Dogs</td>
<td>8.4</td>
<td>99</td>
<td>1992</td>
<td>Crime, Drama, Thriller</td>
</tr>
<tr>
<td>23</td>
<td>Indiana Jones II</td>
<td>7.6</td>
<td>118</td>
<td>1984</td>
<td>Action, Adventure, Fantasy</td>
</tr>
<tr>
<td>27</td>
<td>Die Hard 2</td>
<td>7.1</td>
<td>124</td>
<td>1990</td>
<td>Action, Thriller Crime</td>
</tr>
</tbody>
</table>

**Evaluation:**

- Users in a user study rated movies.
- Analyzed time, year and genre.
- Objective: User would allocate movie 'Die Hard 2' (27) to one of $k=3$ clusters.
- Uses movies (1), (7) and (23) as initial centroids for clusters $C_1$, $C_2$ and $C_3$.

**Tech:**

- Use movies (1), (7) and (23) as initial centroids for clusters $C_1$, $C_2$ and $C_3$.
- Evaluate the squared Euclidean distance.

**Contact:**

info@PreferenceSQL.com
www.PreferenceSQL.com

---

**Application Overview**

The demo application provides a visual comparison of different clustering techniques.

**Benefits are:**

- Movie recommendations which satisfy user preferences.
- Clustering of movies having similar features.
- Exploiting the Borda rule to avoid domain normalization in k-means++. 
- Use quality measures to find suitable values for the desired number $k$ of clusters.

---

**The Borda Social Choice Movie Recommender**

Johannes Kastner,
Nemanja Ranitovic, Markus Endres
Chair for Databases and Information Systems
University of Augsburg, Germany

---

**Contact:**

www.PreferenceSQL.com
info@PreferenceSQL.com