The Borda Social Choice Movie Recommender

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 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 $v_{jm} \in \{0, \dots, k-1\},\ m = 1, \dots, k$, to each candidate. All v_{jm} are pairwise distinct.
1) Sum up the votes for each candidate:

$$bordaSum_{C_i} = \sum_{l=1}^{d} v_{li}$$

*2) Determine the **Borda winner**:*

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

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.

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.

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

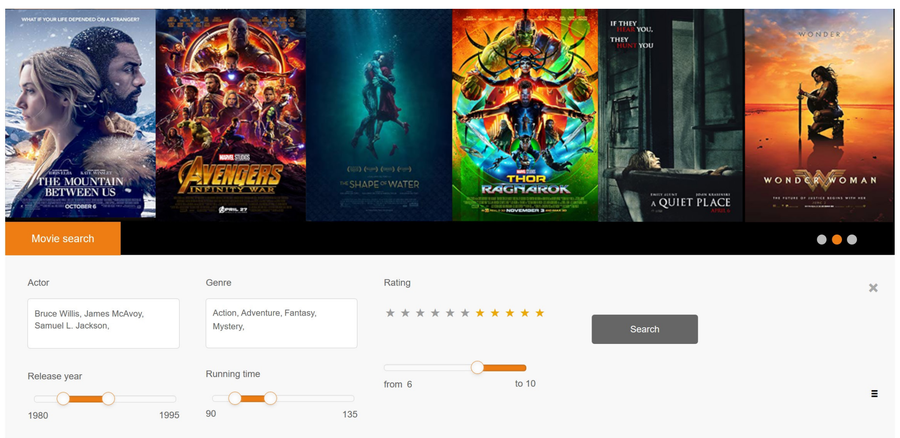
Determine the Borda Winner:

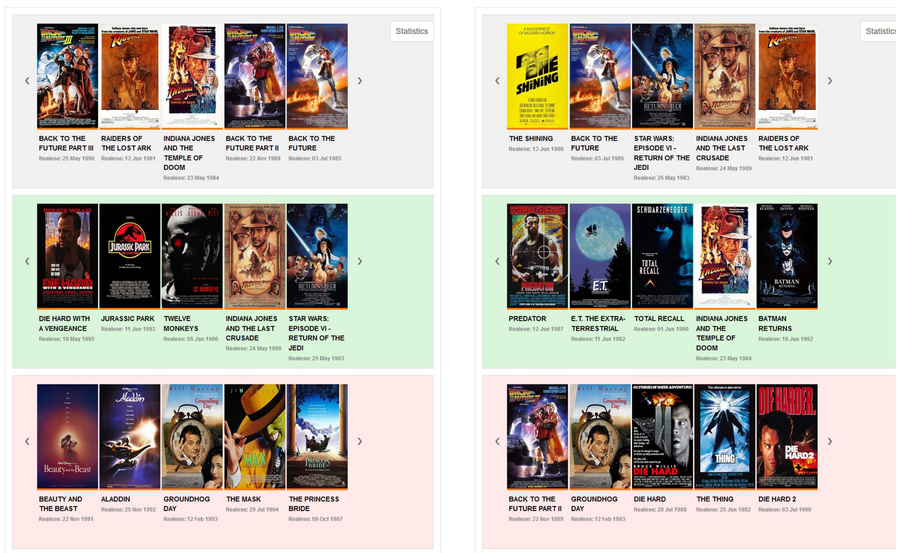
- 'Die Hard 2' is allocated to the Borda Winner C_2.
- Using the squared Euclidean distance, 'Die Hard 2' would be allocated to C_3.

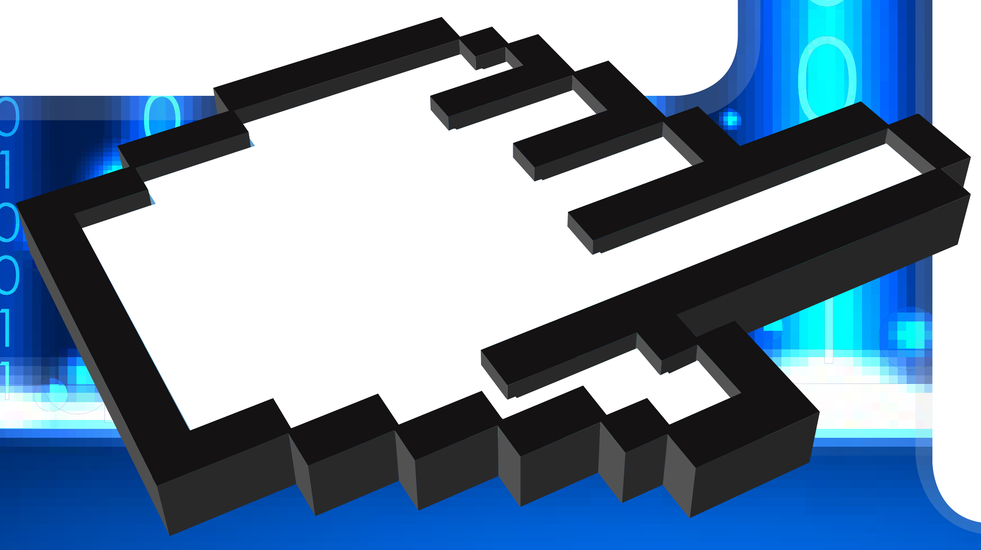
Dimension	C_1	C_2	C_3
rating	1.70 (0)	1.30 (1)	**0.50 (2)**
time	**1.00 (2)**	25.00 (0)	6.00 (1)
year	13.00 (0)	**2.00 (2)**	6.00 (1)
genre	0.25 (1)	**0.50 (2)**	0.20 (0)
Borda Sum	3	**5**	4
Euclidean dist.	13.15	25.12	**8.50**

Demo Architecture:

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







It´s a Preference World

Johannes Kastner,
Nemanja Ranitovic, Markus Endres

Chair for Databases and Information Systems
University of Augsburg, Germany

Universität Augsburg University

Contact:

www.PreferenceSQL.com
info@PreferenceSQL.com