

Content Based Recs

Item profiles

 profile - record or a collection of records representing important characteristics of an item

E.g. Movie

1. a set of actors
2. director
3. writer
4. year of release
5. genre

Where do we get this data?

Non-obvious feature discovery problems:

- documents (TF/IDF, Jaccard sim. — more later)
- images (user tagging, user captioning, object extraction/recognition)
- 3D models (Geometric & Topological shape descriptors)

Rep an item by a profile
each attrib boolean/integral/real valued (can mix)

	chris Columbus	steven Speilburg	George Lucas
Jaws	0	1	0
HPI	1	0	0
HP2	1	0	0
SW1	0	0	1
SW2	0	0	1

More
complex
example
w/ mixed types
of attrib types

	actors								door star
	A1	A2	-	-	-	-	-	A8	
movie A	0	1	1	0	1	1	0	1	3α
movie B	1	1	0	1	0	1	1	0	4α

$$\cos \text{sim}(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

Things to consider:

1. actor in neither movie ... nothing bad

2. # of stars

$$\frac{A \cdot B}{\|A\| \|B\|} = \frac{2 + 12\alpha^2}{\sqrt{25 + 125\alpha^2 + 144\alpha^4}}$$

w/ $\alpha = 1$ $\cos \text{sim}(A, B) = .816$
 $\alpha = \frac{1}{2}$ $\cos \text{sim}(A, B) = .619$
 $\alpha = 2$ $\cos \text{sim}(A, B) = .940$

No right answer

User Profiles

Given Utility matrix

Create vectors (in the same space as items) by user preferences

E.g. Director

$$\begin{matrix} \text{HP} & \{1, 2, 3\} \\ \text{SW} & \{1, 2, 3, 4, 5, 6\} \end{matrix}$$

$$\begin{matrix} \text{Chris} & \text{George} \\ \text{Colombia} & \text{Lucas} \\ 1 & 0 \\ 0 & 1 \end{matrix}$$

Boolean utility matrix

$$\begin{matrix} & \text{HP1} & \text{HP2} & \text{HP3} & \text{SW1} & \text{SW2} & \text{SW3} \\ \text{user A} & 1 & 0 & 0 & 1 & 1 & 0 \end{matrix}$$

 transform to item profile space

$$\begin{matrix} & \text{Chris} & \text{George} \\ \text{user A} & \text{Colombia} & \text{Lucas} \\ 1 & 0 & 1 \end{matrix}$$

Real valued utility matrix

$$\begin{matrix} & \text{HP1} & \text{HP2} & \text{HP3} & \text{SW1} & \text{SW2} & \text{SW3} \\ \text{user B} & 3 & 4 & 2 & 1 & 4 & 3 \end{matrix}$$

 transform to item profile space

$$\begin{matrix} & \text{Chris} & \text{George} \\ \text{user B} & \text{Colombia} & \text{Lucas} \\ \frac{1}{2}[(3-3.5)+(4-3.5)] & & \frac{1}{2}[(1-2.5)+(4-2.5)] \end{matrix}$$

$$\text{for user avg(CC movies)} = 3.5$$

$$\text{avg(GL movies)} = 2.5$$