Rec Systems

Rec using content based method

Find items w/ large sim to a users pref

Given user U
    find item i st $(u,i) = \emptyset$
    w/ "max sim" $(u.prof, i.prof)$

How do we find item w/ "max sim"

1. preprocess interms into some spatial searching
        data structure w/ key $i.prof$
2. find rec:
    a. search data structure for k (approx) close entries   *or maybe a few more*
    b. filter items already of interest
    c. pick most similar from remainder

E.g

RS(u)
    1. u.profile
    2. produce $2k$ close items $\longrightarrow \{i_1, \cdots, i_{2k}\}$

    $sim(u.profile, i_j.prof.le)$ in top $2k$

$\{i_1, i_2, i_3, \cdots\}$
~~HP1~~ HP2 HP3

$a = sim(HP_2.prof, u.prof)$
$b = sim(H_3.prof, u.prof)$
$a > b \longrightarrow \{HP_2, HP_3, \cdots \}$
                        $\underbrace{\qquad\qquad}_{\text{8th item}}$

Collaborative Filering

Don't manually find features

High level
- each user is represented by a row from utility matrix
- recomend for user u
  1. find users $U = \{u_1', \cdots, u_k'\}$ sim to u
  2. recommend item i
     from $u_j' \in U$
     that u_j' is interesete in but $(u, i) = \emptyset$

Measuring sim

	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	4			5	1		
B	5	5	4				
C				2	4	5	
D		3					3

Some examples for considering picking dist
1. A & C   TW & SW1 very diff and should be far apart
2. A & B   HP1 very simmilar (and not much else to go off)

Jaccard Sim

Given 2 sets X Y

$$J_{sim}(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

$$J_{sim_S}(X, Y) = 1 - J_{sim}(X, Y)$$

How to use Jaccard Sim w/ non-set data
  Round!
      - ratings 3 and above are interested
      - ratings below 3 are uninterested