

Personalization and User Modeling

Content-Based Recommender Systems

Patrick Siehndel
{siehndel}@L3S.de

1 User Modeling with Stereotypes

1.1 Solution 1

The following table lists $f_{i,j} = \frac{freq_{i,j}}{\sum_k freq_{k,j}}$ for all pictures / tags

Keyword / Document	pic 1	pic 2	pic 3	pic 4	pic 5
Soccer	$\frac{4}{14}$	$\frac{2}{10}$	$\frac{1}{7}$	$\frac{3}{11}$	$\frac{3}{7}$
Germany	$\frac{3}{14}$	0	$\frac{3}{7}$	0	$\frac{1}{7}$
Portugal	$\frac{3}{14}$	0	0	0	$\frac{1}{7}$
Goal	$\frac{4}{14}$	0	$\frac{1}{7}$	0	0
USA	0	$\frac{4}{10}$	0	0	$\frac{1}{7}$
Fans	0	$\frac{4}{10}$	$\frac{2}{7}$	0	0
Team	0	0	0	$\frac{4}{11}$	0
Ghana	0	0	0	$\frac{4}{11}$	$\frac{1}{7}$

Next we calculate $idf_i = \log \frac{D}{d_i}$

Keyword	idf_i
Soccer	$\log \frac{5}{5}$
Germany	$\log \frac{5}{3}$
Portugal	$\log \frac{5}{2}$
Goal	$\log \frac{5}{2}$
USA	$\log \frac{5}{2}$
Fans	$\log \frac{5}{2}$
Team	$\log \frac{5}{1}$
Ghana	$\log \frac{5}{2}$

The resulting weighted term vectors are the following:

$$picture_1 = \begin{pmatrix} 0 \\ 0.0475 \\ 0.0853 \\ 0.1137 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$picture_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0.1592 \\ 0.1592 \\ 0 \\ 0 \end{pmatrix}$$

$$picture_3 = \begin{pmatrix} 0 \\ 0.0951 \\ 0 \\ 0.0568 \\ 0 \\ 0.1137 \\ 0 \\ 0 \end{pmatrix}$$

$$picture_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.2542 \\ 0.1447 \end{pmatrix}$$

$$picture_5 = \begin{pmatrix} 0 \\ 0.0317 \\ 0.0568 \\ 0 \\ 0.0568 \\ 0 \\ 0 \\ 0.0568 \end{pmatrix}$$

1.2 Solution 2

For calculating the cosine similarity we consider all dimensions of the word vectors, not only the ones the compared pictures have in common.

We calculate the cosine similarity between picture 1 and picture 3

$$picture_1 = \begin{pmatrix} 0 \\ 0.0475 \\ 0.0853 \\ 0.1137 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad picture_3 = \begin{pmatrix} 0 \\ 0.0951 \\ 0 \\ 0.0568 \\ 0 \\ 0.1137 \\ 0 \\ 0 \end{pmatrix}$$

$$cossim(p1, p3) = 0.46135$$

next we calculate the cosine similarity between picture 2 and picture 3

$$picture_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0.1592 \\ 0.1592 \\ 0 \\ 0 \end{pmatrix} \quad picture_3 = \begin{pmatrix} 0 \\ 0.0951 \\ 0 \\ 0.0568 \\ 0 \\ 0.1137 \\ 0 \\ 0 \end{pmatrix}$$

$$cossim(p1, p4) = 0,50648$$

1.3 Solution 3

We define a very simple vector for a user who likes “Goals” and “Portugal”.

$$u = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{When comparing this vector with our 5 pictures we see that}$$

pictures 2 and 5 are not related to “Goals” or “Portugal”. We now calculate the cosine similarity between the user vector and the remaining three pictures.

$$cossim(u, p1) = 0,6265$$

$$cossim(u, p3) = 0,253$$

$$cossim(u, p5) = 0.3886$$

If we would calculate the cosine similarity only based on the tags the pictures have in common we would get a similarity of 1 for p3 and p5