

PCA

k - NN

## Humor Document Recommendation

using Adaptive k-NN with PCA

,

Jongwoo Lee, Byung - Tak Zhang

School of Computer Science and Engineering, Seoul National University

{jwlee, btzhang}@scai.snu.ac.kr

(preference)

MrHumor

. MrHumor

(vector space)

PCA

2

k - NN

.

### ABSTRACT

We constructed *MrHumor* System which analyzes the preferences of users on internet, learning collaborative filtering technique, and recommends unrated humor documents to the users. In *MuHumor*, the vector space of rating vectors of users on humor documents is defined and transformed to 2-dimensional space by PCA to be immune to noises and to operate effectively, where the preferences of new documents are predicted using adaptive k-NN with statistical characteristics of humor documents.

1. [Pazzani, 1999], [Lin, 2000] (information recommendation), (information filtering) (locality)

가

(demographic), (contents filtering), PCA (collaborative filtering) (demographic) (profile) (item)

II.

가

$$R_u = (r_{u1}, r_{u2}, \dots, r_{uN})$$

$$R = \{R_1, R_2, \dots, R_U\}$$

$$r_{ui} \quad u \quad i$$

가

가

$$R_i \quad R_j \quad k \quad , \quad r_{ik}$$

$$r_{jk} \quad N-1$$

$$|r_{ik} - r_{jk}|$$

가

가

(Pearson correlation)

MrHumor

[Resnick, 1994][Shardanand, 1995], [Gupta, 1998], [Chien, 1998], 가 가 가

$|R_i - R_j|$   
 가  
 가  
 MrHumor  
 $R_u$   
 PCA (Principal Component Analysis)  
 $r_{ui}$   $u$   
 k  
 k  
 2  
 47  
 80  
 3760  
 가  
 가  
 20  
 94  
 가  
 1)  
 2)  
 NN  
 MSE

**1. k-NN**  
 $r_u = (r_{u1}, r_{u2}, \dots, r_{uN})$   
 $R = \{r_1, r_2, \dots, r_U\}$   
 1.  

$$z_u = \left( \frac{r_{u1} - \text{Exp}(r_u)}{\sigma(r_u)}, \dots, \frac{r_{uN} - \text{Exp}(r_u)}{\sigma(r_u)} \right)$$
  
 2. **Z** SVD(Singular Value Decomposition)  

$$Z = U \Sigma V^T$$
  
 3. 가 가 2  
 $V$   
 4. 2  
 $Y = \{y \mid y = zV^i, z \in Z\}$   
 5.  $K_{iu}$   $u, i$   
 $K = \{(K_{iu}, \sigma_i^2)\}$  (KNN(K,u)  $y$   
 $u$  가 가  $K$   
 .)  

$$K_{iu} = \arg \min_K \left( \frac{\sum_{j \in \text{KNN}(K,u)} r_{ji}}{K} - r_{ui} \right)$$
  

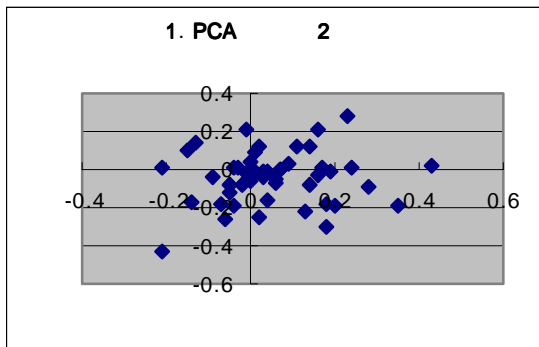
$$\sigma_i^2 = \text{VAR}_{\forall u}(r_{ui})$$
  
 6. 5  $K$   $K_{iu}$   $\sigma_i^2$

PCA  
 2  
 가 가  
 3) Cluster  
 2 8  
 4)  
 k-NN  
 2 가 가  $K$

. K 1.  
**K**  
 1  
 1. 2), 3), 4)  
 PCA 47  
 2

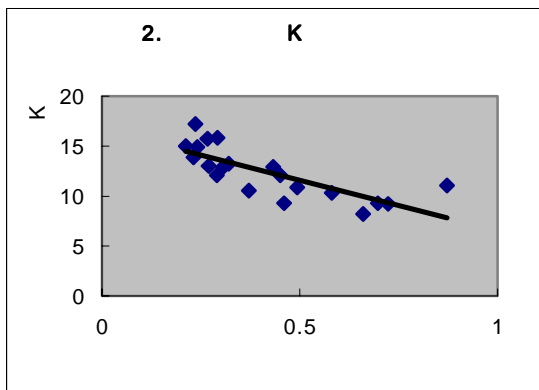
4가

, NN,



k - NN

(AITRC)



1. 5, 6 **K**  
 2.  
 V K

$$K = -10.1V + 16.6 \quad (1)$$

1. 4가

	MSE			
		NN		k - NN
MSE	0.583	0.899	0.539	0.421

1.

[Resnick, 1994] P. Resnick and N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An open architecture for collaborative filtering," *Proc. ACM 1994 Conf. on Computer Supported Cooperative Work*, pp. 175-180.

[Gupta et al., 1999] D. Gupta, M. DiGiovanni, H. Narita, and K. Goldberg, "Jester2.0: Evaluation of a new linear time collaborative filtering algorithm applied to jokes," *Poster Session and Demonstration, 22nd Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 191-192.

[Chien, 1998] Y. Chien, "A Bayesian model for collaborative filtering," *Proc. of Uncertainty in Artificial Intelligence*, Morgan Kaufmann.

[Billsus and Pazzani, 1999] D. Billsus and Michael J. Pazzani, "Learning Collaborative Information Filters," *In Shavlik, J., ed., Machine Learning: Proc. of the Fifteenth International Conference*, Morgan Kaufmann Publishers.

[Lin, 2000] W. Lin, S. A. Alvarez, C. Ruiz, "Collaborative Recommendation via Adaptive Association Rule Mining," *Workshop on Web Mining for E-Commerce*, pp. 35-41.

[Shardanand, 1995] U. Shardanand and P. Maes, "Social Information Filtering: Algorithms for Automating Word of Mouth", *Proc. ACM CHI95*, pp.210-217.