http://dx.doi.org/10.7236/JIWIT.2012.12.2.61
JIWIT 2012-2-8

시간 정보를 이용한 확장성 있는 하이브리드 Recommender 시스템

Scalable Hybrid Recommender System with Temporal Information

Farman Ullah, Ghulam Sarwar, 김재우*, 문경덕**, 김진태**, 이성창***

Farman Ullah, Ghulam Sarwar, Jae-Woo Kim, Kyeong-Deok Moon, Jin-Tae Kim, Sung-Chang Lee

요 약 최근 디지털 컨텐츠와 컨텐츠 사용자의 기하 급수적인 증가와 함께 recommender 시스템이 주목을 받으며 많은 응용 프로그램에 적용되고 있는 가운데, recommender 시스템의 확장성과 대체적으로 이와 반비례하는 정확성이 이슈가 되고 있다. 본 논문에서는 recommender 시스템 모델 중 하이브리드 모델의 매트릭스를 제거하고 아이템의 특성을 정하기 위해 클러스터링 기술을 사용한 Scalable Hybrid Recommender System을 제안한다. 제안된 모델은 recommender 시스템의 확장성과 정확성을 향상시키기 위해서 아이템에 대한 사용자의 평가 정보, demographic 정보와 구체적인 시간 정보를 사용한다. Reduction 기술 사용을 통해 Item-feature 매트릭스의 사이즈를 축소하고, 사용자 demographic 정보를 사용하여 temporal aware hybrid user model을 만든 후, 비슷한 정보를 가진 사용자간 클러스터링을 통해, 가장 유사한 정보를 가진 사용자들을 추출하여, 사용자간 정보를 비교함으로써 사용자가 원하는 아이템의 특성을 예상하고 사용자에게 N개의 아이템을 추천함으로써, 기존의 recommender 시스템보다 더욱 향상된 결과를 도출해 낼 수 있는 알고리즘을 제시하였다.

Abstract Recommender Systems have gained much popularity among researchers and is applied in a number of applications. The exponential growth of users and products poses some key challenges for recommender systems. Recommender Systems mostly suffer from scalability and accuracy. The accuracy of Recommender system is somehow inversely proportional to its scalability. In this paper we proposed a Context Aware Hybrid Recommender System using matrix reduction for Hybrid model and clustering technique for predication of item features. In our approach we used user item-feature rating, User Demographic information and context information i.e. specific time and day to improve scalability and accuracy. Our Algorithm produce better results because we reduce the dimension of items features matrix by using different reduction techniques and use user demographic information, construct context aware hybrid user model, cluster the similar user offline, find the nearest neighbors, predict the item features and recommend the Top N- items.

Key Words : Item features rating, User Demographic information, Contextual information, Matrix Reduction Techniques, k-means clustering Algorithm.

*준회원, 한국항공대학교 통신공학과
**준회원, 한국전자통신연구원 그린컴퓨팅연구부
***정회원, 한국항공대학교 항공전자 및 정보통신공학부(교신저자)
접수일자 2012.4.4, 수정일자 2012.4.12
게재확정일자 2012.4.13

Received: 4 April, 2012, Revised: 12 April, 2012
Accepted: 13 April, 2012
***Corresponding Author: sclee@kau.ac.kr
Dept. of Information and Communication Engineering, Korea Aerospace University, Korea

Ⅰ. INTRODUCTION

User Preference and Recommendation Systems have gained much popularity and are widely used in E - Commerce, Automatic IPTV content recommendation, Media sharing, Net Surfing and web service selection etc. Choosing among a number of items is challenging job for a user. Advancement in Technology and growing density of contents and users demands recommender system to make an accurate recommendation in user bearable time. Most of recommender systems use Collaborative filtering[1, 2, 3] that work on the idea that similar users have similar preferences. Collaborative filtering can be divided into two types, memory-based[4] and model-based[5].

In this paper, we used two different matrix reduction techniques i.e. Singular Value Decomposition and Relative Feature Rating in specific time and day to reduce the dimension of items feature rating matrix. Then we construct Temporal Aware Hybrid User Model from the reduced matrix and User Demographic information. We applied K-means Clustering Algorithm [6] on the Hybrid Model, in order to cluster the similar users offline. Do neighbors selections from the cluster to which user belong to reduce online processing time. The final recommendation is based on the contextual information, neighbors set and rating matrix. Keeping temporal information in term of Day, Time improves the accuracy of the system. We predict the item features and select the Top N-items from the rating of the neighbors set.

The paper is organized as follow: Section-II briefly describe about the matrix reduction techniques and its calculation. Section-III describe the Temporal Aware Hybrid Recommender System, Section-IV Temporal Hybrid User Model, Section-V clustering technique, Section-VI is about neighbor selection and predication of item features & item selection, Section-VII is about our simulation and results and the last section about conclusions and future work.

Ⅱ. MATRIX REDUCTION TECHNIQUES

In order to improve the scalability of recommender system we reduce the size of the items features rating matrix. We use two different techniques to reduce the matrix. Using the statistical measure of Mean Absolute Error (MAE), due to less MAE we recommend relative feature rate for reduction of matrix as shown from Fig. 1.

1. Relative Feature Rating

In Relative Feature Rating Technique we find the relative mean of each feature by using the item feature rating that user rate at specific time and day. The related concepts and definition we used are as under:-

Definition 1 Total Feature Rating: the sum of ratings R that user 'i' grades for all the item attributes at a specific time and day, which is expressed as TFR(i,t,d).

$$\mathrm{TFR}\,(i,t,d) = \textstyle\sum_{j\in Ti} \mathrm{R}(i,j,t,d) \qquad (1)$$

Here Ti stands for the set composed of items have been graded by user i.

Definition 2 Feature Rating: the sum of ratings that user i grades for the feature k in that time and day, which is expressed as FR(i,k,t,d).

$$\mathrm{FR}\,(i,k,t,d) = \sum_{k\in Fk\subset Ti} \mathrm{R}(i,k,t,d) \qquad (2)$$

Definition 3 Relative Feature Rating: relative feature rating that user i grades for feature k, expressed as RFR, is defined as the ratio of Feature Rating FR(i,k,t,d) and Total Feature Rating TFR(i,t,d).

$$\mathrm{RFR}\,(i,k,t,d) = \left(\frac{\mathrm{FR}(i,k,t,d)}{\mathrm{TFR}(i,t,d)}\right) * \alpha \qquad (3)$$

Since the relative is normally a value less than one, so we use is "α" a constant factor in our algorithm. In our algorithm we use seven days a week and four time ranges in each day. In order to understand above mentioned concepts we consider five different movies that are rated by three users at four time ranges at Day

표 1. 사용자에 의해 평가된 영화/아이템 특성표
Table 1. Movies/Items features rating by user at time Ti and day Di

Movies/items	User1				User2				User3				Context Information	
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	Time	Day
1	4	5	5	4	5	3	1	2	2	2	3	2	t1	d1
2	4	3	2	3	4	2	3	3	4	4	5	1	t1	d1
3	2	1	4	3	3	5	3	4	3	3	4	4	t1	d1
4	5	4	3	5	4	4	5	1	1	4	3	3	t1	d1
5	1	2	4	2	4	3	2	2	3	2	5	1	t1	d1
1	0	4	5	4	4	4	2	5	1	4	2	2	t2	d1
2	4	0	2	5	2	2	4	5	4	2	3	1	t2	d1
3	2	3	3	2	2	1	1	1	2	5	3	2	t2	d1
4	2	3	3	4	3	2	2	1	4	5	4	4	t2	d1
5	3	4	4	3	4	2	3	2	3	5	2	2	t2	d1
...	..	..	.	..	..	..	..	..	..	..	..	..	..	..
1	1	2	2	2	3	2	2	3	2	5	3	2	t1	d2
2	4	5	3	1	5	3	4	3	3	5	4	1	t1	d2
3	2	1	4	2	2	3	2	4	4	4	5	2	t1	d2
4	0	4	3	3	4	2	3	2	3	3	4	3	t1	d2
5	1	3	2	3	2	4	3	3	3	2	3	3	t1	d2
…………														
1	3	2	1	3	3	5	3	2	4	1	3	5	t1	d3
2	5	2	2	3	2	3	3	2	4	2	3	5	t1	d3
3	4	2	3	2	3	4	2	3	4	2	3	5	t1	d3
4	5	3	4	3	4	2	1	4	4	3	3	5	t1	d3
5	4	3	4	4	5	5	3	3	4	4	3	5	t1	d3
…………														
1	0	4	3	3	4	2	3	2	5	4	3	3	t1	d6
....	....	....	...	...	...	...	..	...	....	...	...	...	...	...
5	1	3	2	3	2	4	3	3	2	4	2	3	t4	d7

d1-d7. The Table.1 shows the data we use for understanding of above concepts. Table.2 shows the reduced matrix after applying the Equation.1, 2 and 3.

표 2. 사용자를 위한 RFR
Table 2. RFR for users at time Ti and day Di

Users	Relative Feature Rating				Context Info.	
	C1	C2	C3	C4	Ti	Di
U1	2.42	2.27	2.73	2.58	T1	D1
U2	3.17	2.7	2.22	1.9	T1	D1
U3	2.2	2.54	3.39	1.86	T1	D1
U1	1.83	2.33	2.83	3	T2	D1
U2	2.11	2.5	2.24	3.16	T2	D1
U3	3.33	2.27	2.88	1.5	T2	D1
U1	1.67	3.13	2.92	2.29	T1	D2
U2	2.71	2.37	2.37	2.54	T1	D2
U3	2.34	2.97	2.97	1.72	T1	D2
U1	3.39	1.94	2.26	2.42	T1	D3
U2	2.74	3.06	1.94	2.26	T1	D3
U3	2.78	1.67	2.08	3.47	T1	D3

2. Singular Value Decomposition

Singular Value Decomposition (SVD), is a matrix factorization technique[7,8] that decomposes a single matrix into three different matrices.

$$\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^{*} \qquad (4)$$

If M is a matrix of m x n, then U is m x m an unitary matrix, Σ is an m×n rectangular diagonal matrix, and V∗ (the conjugate transpose of V) is n×n an unitary matrix. The diagonal entries $\Sigma_{i,i}$ of Σ are known as the singular values of M [8].

We apply the SVD in order to find the singular value for each feature rated by a user at specific time and day. We normalize the singular values matrix by dividing it the maximum singular value and multiply it by a constant, so the mean value in RFR and SVD is 2.5 in our algorithm. So we find the Mean absolute error of both reduction methods by using different number of users that rate the movies at specific time and day as shown in Fig. 1.

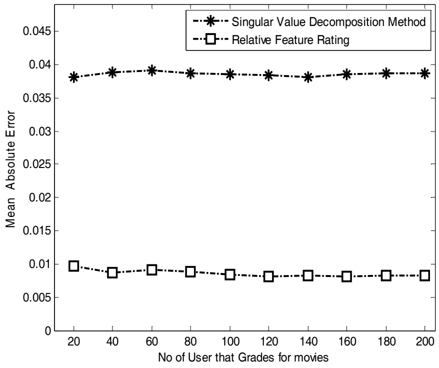

그림 1. RFR과 SVD를 사용한 매트릭스 축소
Fig. 1. Matrix reduction using RFR and SVD

III. Scalable hybrid recommender system with temporal information

As the number of users and items are growing exponentially, so it becomes difficult for recommender system to make an accurate recommendation in user bearable time. Some Conventional Recommender system uses only item rating for recommendation [9], some improved recommender system uses attribute rating[10] and demographic similarity[11]. The conventional recommender system does not consider any temporal information that whether user a like the item at that specific time or not, it find the similarity of item with all database, which reduce the scalability and increase the recommendation time.

The Workflow of our algorithm i.e. scalable hybrid recommender system with temporal information is shown in Fig.2. In our algorithm we use the temporal information that is four time ranges per day and seven days a week. We use the rating and demographic information in order to predict the feature value if a user have not assigned to any attribute in order to eliminate the sparsity problem. When we dense the rating matrix, we apply the matrix reduction technique i.e. Relative Feature Rating in order to reduce the dimension of the rating matrix. We construct the temporal aware hybrid model by using the relative feature rating matrix, user demographic information and context information i.e. time and day. We apply the k-means clustering algorithm to temporal aware hybrid user model in order to cluster the similar users. The process from feature rating to clustering the similar users is offline. We find the target user belong to which cluster, then similarity with users and select the Top N- similar neighbors. Using the similar neighbor we predict the item features and recommend the Top N-similar items.

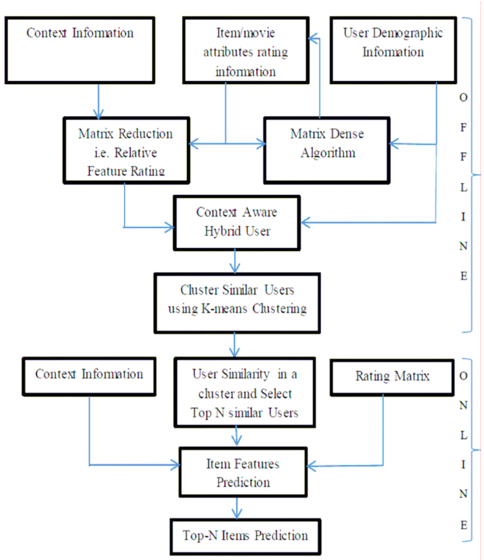

그림 2. 시간 정보를 이용한 Scalable Hybrid Recommender 시스템의 워크플로우
Fig. 2. Workflow of Scalable Hybrid Recommender system with temporal information

IV. Temporal aware hybrid User model

We construct the temporal aware hybrid user model offline by using relative feature rating matrix, user demographic information and temporal information. We have considered four demographic features of each user. The temporal aware hybrid user model is a three dimensional matrix as shown in Table III. The temporal aware hybrid user model has the following advantages:-

Ø The temporal aware hybrid user model has a reduced dimension as compared to the items feature rating matrix, so it reduces the computational complexity.

Ø We apply the clustering technique on this model offline to cluster the similar users so that the system recommends an accurate item with user bearable time.

표 3. 상황 인식 하이브리드 사용자 모델
Table 3. Context aware hybrid user model

User	Relative Features Rating			User Demographic Features			Context Info.	
	RFR1	...	RFRp	D1	.	Dq	Ti	Di
U1	H11	...	H1p	H1(p+1)	.	H1(p+q)	T1	D1
...	...	...	...	...	.	...		
Um	Hm	...	Hmp	Hm(p+1)	.	Hm(p+q)	T1	D1
U1	H11	...	H1p	H1(p+1)	.	H1(p+q)	T2	D1
...	...	...	...	...	.	...	..	
Um	Hm	...	Hmp	Hm(p+1)	.	Hm(p+q)	T2	D1
..	..	..	..	...	.	..	..	..
Um	Hm	...	Hmp	Hm(p+1)	.	Hm(p+q)	T4	T7

V. Clustering of Simmilar users

Clustering means to divide the items into similar groups based on some similarities. Cluster analysis could be divided into hierarchical clustering and non-hierarchical clustering techniques. Examples of hierarchical techniques are single linkage, complete linkage, average linkage, median, and Ward. Non-hierarchical techniques include k-means, adaptive k-means, k-medoids, and fuzzy clustering[12, 13]. A good clustering algorithm is one who produces groups with non-overlapping. We apply the k-means clustering technique to temporal aware hybrid user model to group the similar users into k-groups. In our algorithm we use the cosine similarity in order to group the similar users. The algorithm is explained in Fig.3 and Fig.4.

Input:- Matrix of Context Aware Hybrid User Model of size m rows and n columns for each time and day

Output:- A vector having the Group i.e. cluster number of mx1, and a matrix having the centroid for each cluster of size k x n.

Steps:-

1. *Select k-cluster with initial centroids*
2. *Partition the matrix into k-clusters(C_1, C_2 ...C_k)* based on cosine similarity in such a way that $C_1 \cap C_2 = ø$ and $C1 \cup C2 \cup \ldots C_k$ *is the original matrix*

그림 3. Cosine similarity를 이용한 K-means 클러스터링
Fig. 3. K-means clustering using cosine similarity

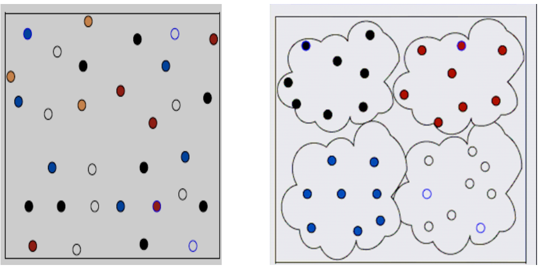

그림 4. 그룹 내 유사한 정보를 가진 사용자 클러스터링
Fig. 4. Clustering similar users in a group

Ⅵ. Neighbors selection and item features predication and item selection

The process in our algorithm from items feature rating to clustering similar users in a group is offline, which improve the performance of recommender system. We find the Top N-similar users in a group/cluster at a specific time and day not in a whole database of items or hybrid model. We find the minimum absolute error (MAE) using Eq.5 of target user with the centroid of each group.

$$MAE = \frac{\sum_{k=1}^{n} |p_k - q_k|}{n} \tag{5}$$

We select the cluster that having minimum MAE with the target user and using relative cosine similarity Eq. 6 to find the Top N most similar user in that group at that specific time and day.

$$\mathrm{sim}(a,b,t,d) = \frac{\sum_{k=1}^{p+q} |(\mathrm{Haktd} - \overline{\mathrm{Hatd}})\,\mathrm{X}\,(\mathrm{Hbktd} - \overline{\mathrm{Hbtd}})|}{\sqrt{\sum_{k=1}^{p+q} (\mathrm{Haktd} - \overline{\mathrm{Hatd}})^2\,\mathrm{X}\,\sum_{k=1}^{p+q}(\mathrm{Hbktd} - \overline{\mathrm{Hbtd}})^2}} \quad (6)$$

Where Haktd and Hbktd are the relative interest score for feature k in time ‘t’ and day ‘d’ for target user a and user b from cluster. Hatd and Hbtd is the average rating of User a and b in time ‘t’ and day ‘d’ respectively.

1. Items Feature Predications and Item Selection Based on Similar Users

After obtaining the similar users from a cluster with the target user, we predicate the target user feature rating by using Eq.7.

$$P(a,j,t,d) = \left\{ R_{atd} + \frac{\sum_{i\in U} \mathrm{Sim(a,i,t,d)\ X\ (Rijtd} - \overline{\mathrm{Ritd}})}{\sum_{i\in U} \mathrm{sim(a,i,t,d)}} \right\} \quad (7)$$

Where P(a,j,t,d) is the predicted rating that user a grades for feature j at time t, and U is the nearest neighbor set of user "a" calculated at that time and day. Rijtd is rating of neighbor i for the content j in that time and day. Ratd is the rating to user a in time t and day d. After obtaining the predicated item features, we select the Top N- most similar items from the similar neighbor set we selected for feature predication.

VII. Simulation and Results

Currently no such movies or items database we find that has rating at specific time and day. So we build our own database having 1000 movies rated by 200 users at four time ranges per day and seven days a week. Each movie has four feature rating and user has four demographic information. We compare our algorithm with Collaborative Filtering using Hybrid User Model[14], we consider the 1000 movies database for the Collaborative Filtering using Hybrid User Model also.

We use 80% of the data as a training data and 20% for testing. Many statistical measuring can be used, but we use the mean absolute error of Eq.5 and performance i.e. recommendation time. Fig.5 shows that there is a little difference in mean absolute error; the collaborative filtering using hybrid user model has less error as compared to our algorithm, because we consider a cluster of users not the whole users matrix. But there is a tremendous difference in the recommendation time as shown in Fig.6.

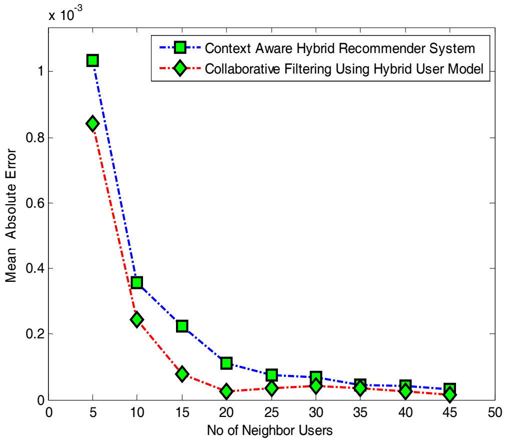

그림 5. 예상 오류 : 하이브리드 사용자 모델과 시간정보를 사용한 확장성 있는 하이브리드 사용자 모델을 사용한 Collaborative 필터링
Fig. 5. Predication error : Collaboratice filtering using hybrid user model and Scalable Hybrid Recommender System with Temporal Information

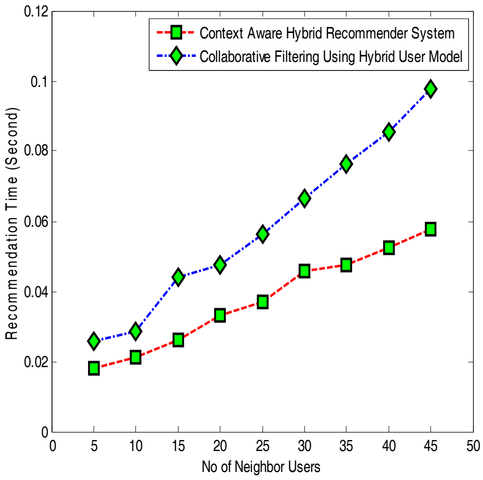

그림 6. 추천 시간
Fig. 6. Recommendation time

VIII. CONCLUSION AND FUTURE WORK

The traditional recommender systems do not consider the context, so due to increasing the items and users it becomes difficult to recommend an accurate item with user bearable time. Our algorithm solves the problem of sparsity, scalability and cold start issues. We apply our algorithm for recommendation of movies, but it can be used anywhere for recommendation.

We have considered only the time and day as context information, but we will improve our recommender system by adding further context information and user demographic information. We are also working by adding the device information for N screen service architecture.

참 고 문 헌

[1] Herlocker, J., Konstan, J., Borchers, A., and Riedl, J. (1999). An Algorithmic Framework for Performing Collaborative Filtering. In Proceedings of ACMSIGIR'99. ACMpress.

[2] Konstan, J., Miller, B., Maltz, D., Herlocker, J.,Gordon, L., and Riedl, J. (1997). GroupLens:A pplying Collaborative Filtering to Usenet News. Communications of the ACM, 40(3), pp. 77-87.

[3] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). GroupLens:A n Open Architecturefor Collaborative Filtering of Netnews. In ProceedingsofCSCW'94,ChapelHill,NC.

[4] Shardanand, U., & Maes, P. (1995). Social information filtering: Algorithms for automating 'Word of Mouth'. In proceedings of the conference on human factors in computing systems,

[5] Shahabi, C., Banaei-Kashani, F. Chen, Y., & Mcleod, D. (2001). Yoda: accurate and scalable web-based recommendation systems, In the preceeding of the sixth international conference on cooperative information systems (coopIS 2001), Trento Italy.

[6] Badrul M. Sarwar, George Karypis, Joseph Konstan, John Ried: Recommender Systems for Large-Scale E-Commerce: Scalable Neighborhood Formation using Clustering, the fifth international Conference on Computer and Information Technology ICCIT 2002.

[7] http://www.cs.wits.ac.za/~michael/SVDTut.pdf

[8] http://en.wikipedia.org/wiki/Singular_value_decomposition

[9] "PuWANG", "HongWu YE" A Personalized Recommendation Algorithm combining Slope One Scheme and User Based Collaborative Filtering, 2009 International Conference on Industrial and Information Systems

[10] "HengSong Tan", "HongWu Ye" A Collaborative Filtering Algorithm Based on Item Classification, 2009 Pacific-Asia Conference on Circuits and Communications and System

[11] "YaE Dai", "Hong Wu Ye", Personalized Recommendation Algorithm using User Demographic Information, Second International Workshop on Knowledge Discovery and Data Mining.

[12] "Oyelede, O. J", "Oladipupo, O. O", " Obagbuwa. I. C", Application of k-Means Clustering algorithm for prediction of Students' Academics Performane, International Journal of Computer Science and Information Security, Vol. 7, N0. 1, 2010

[13] "Dong-Moon Kim", "Kun-su Kim", "Kyo-Hyun Park", " Jee-Hyong Lee", " Keon Myung Lee", A Music Recommendation System with a Dynami K-means Clusting Algorithm, Sixth Internation Conference on Machine Learning and Applications

[14] "Qain Wang", "Xianhu Yuan", "Min Sun", Collaborative Filtering Recommendation Algorithm based on Hybrid User Model, 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery.

※ This work was supported by the Industrial Strategic technology development program(10039202, Development of SmartTV Device Collaborated Open Middleware and Remote User Interface Technology for N-Screen Service) funded by the Ministry of Knowledge Economy(MKE,Korea)"

저자 소개

Farman Ullah(비회원)

- 2002-2006: B.Sc Computer Systems Engineering, UET Peshawar, Pakistan
- 2008-2011: M.Sc Computer Engineering, Center For Advanced Studies in Engineering, Pakistan

Ghulam Sarwar(비회원)

- 2003-2007: B.Sc Computer Systems Engininеering, UET Peshawar, Pakistan.

김 재 우(비회원)

- 2003-2011:한국항공대학교 통신공학과(공학사)

문 경 덕(비회원)

- 1990: 한양대학교 전자계산학과(공학사)
- 1992: 한양대학교 전자계산학과(이학석사)
- 2005: ICU 대학원 정보통신공학부(공학박사)
- 1992-1996: 시스템공학연구소 연구원
- 1997-1999: 한국전자통신연구원 선임연구원
- 2000- 현재:한국전자통신연구원 그린컴퓨팅미들웨어연구팀장

<주관심분야: 홈네트워크, u-컴퓨팅, Green ICT, IT조선융합>

김 진 태(비회원)

- 1980년: 인하대학교 전자공학과(공학사)
- 1982년: 인하대학교 전자공학과 (공학석사)
- 1996년: 인하대학교 전자공학과 (공학박사)
- 1987-1988: University of Missouri Kansas City 방문연구원
- 1979-현재: 한국전자통신연구원 융합기술연구부문 그린컴퓨팅연구부 책임연구원

<주관심분야: 스마트홈네트워크, 스마트TV, N-스크린>

이 성 창(정회원)

- 1976-1980:경북대학교 전자공학과(공학사)
- 1983-1985:한국과학기술원 전기 및 전자공학과(공학석사)
- 1987-1991:Texas A&M University 전기전자공학과(공학박사)
- 1985-1987:한국과학기술원 시스템공학센터
- 1992-1993:한국전자통신연구원
- 1993-현재:한국항공대학교 항공전자 및 정보통신공학부 교수

<주관심분야: 방송통신융합, 스마트네트워크>