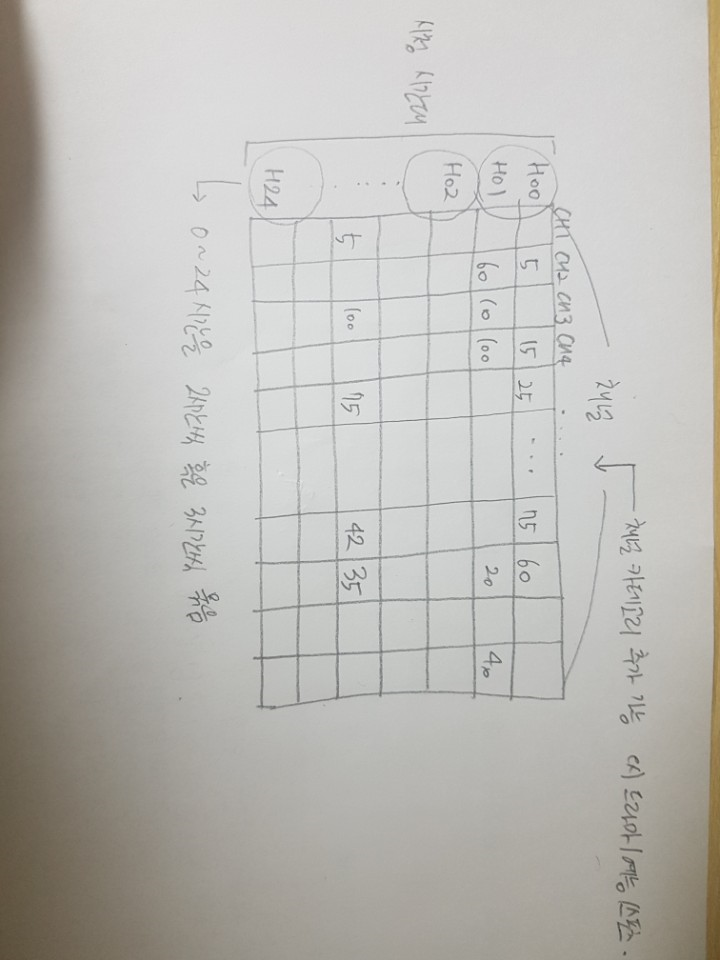
**TV 시청 패턴을 고려한 사용자 분류**

Idea ) 1. 한 가정내에서 시간대에 따른 feature vector 뽑아내 clustering한다.

(50가구 30일치 데이터 요청)

2. 클러스터링된 결과 나타나는 패턴을 레이블로 삼아 가정 내에서 classification

3. Training data/Test data의 비율은 80:20 정도

4. 향후, co-watching matrix도 적용해보면 좋을 것

feature vector 형태 :

* 각 채널을 시청한 시간(분)이 value로 들어감
* 시간대를 일정하게 묶어 (2-3시간 정도) 하나의 vector로 만든다.
* 채널 카테고리도 추가 ex) 드라마/예능/스포츠
* 시간대별로 묶이므로 Data sparsity 어느정도 해결 가능
* 만들어진 feature vector를 clustering한다.
* 이때, objective function은 within-cluster variance를 최소화하는 방향
* K-means objective function과 동일 2
* 실험의 objective function이 K-means objective function과 동일함을 보이며 클러스터링을 통해 이 실험을 하게 되는 의미를 밝히는 것이 중요

추가 제안: co-watching matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 0 |  |  |  |  |
|  |  |  | 1 |  |  |
|  |  |  |  |  |  |
|  | 1 |  |  |  |  |
|  |  | 1 |  |  |  |
|  |  |  |  |  |  |

채널

채널

가로/세로를 모두 채널로 하는 matrix를 만들어 같은 시간대에 틀어진 채널이 있다면 1로, 그렇지 않다면 0의 value를 갖도록 하는 matrix를 만든다. Ex) 2번 채널과 4번 채널이 같은 시간대에 틀어졌다면 (2,4) = 1로 표시

* 효과: 같은 시간대/같은 채널 카테고리로 합쳐져 무시되던 채널이 co-watching matrix를 이용하면 구분될 수 있다.