1. **Rating estimation in Multidimensional Recommender Systems overview** (Chapter 4 in the “key” paper)

* Problem: given the initial (small) set of user-assigned ratings specified for different levels of the multidimensional cube of ratings, the task is to estimate all other ratings in the cube at all the levels of the OLAP hierarchies.
* Not all the traditional 2D Recommender System can be directly extended to the multidimensional case because extra dimensions, profiles, and aggregation hierarchies complicate the problem
* **Aggregation Hierarchy (page 19)**:

Eg: R(JD, action) = 6 and that R(JD, Gladiator) = 7 and R(JD, Matrix) = 3. Then how can we use the cumulative rating of action movies R(JD, action) = 6 to estimate ratings of other individual action movies that John has not seen yet?

🡪 Very complex problem, outside the scope of this paper (and our thesis)

* **Comprehensive Profiles**. In Section 3.2 we described comprehensive profiling methods which have not been used in any of the recommendation methods described in Section 2. We can use this profiling information not only to provide richer recommendations as

mentioned in Section 3.2, but also for rating estimation purposes. For instance, given

rule-based profiles of users, the recommender system can use the discovered rules to

reduce the recommendation space as demonstrated in the following example. Assume

that John Doe’s profile contains the rule “IF time-of-week=weekend THEN movie-type=

action.” Then the recommender system might use only action movie ratings to provide

movie recommendations to John Doe on weekends.

🡪 Sure we will not consider this too =]

* **Muliple Dimensions**

some of the two-dimensional techniques can be directly extended to the multidimensional case. In addition, we propose to consider the reduction-based estimation method

🡪 There are 3 multidimensional rating estimation approaches: Reduction-based, Heuristic-based(Memory-based), Model-based.

🡺 We focus on one special case of the multi-level multidimensional rating estimation problem having **no OLAP hierarchies, where only simple attribute based profiles are used and only individual multidimensional ratings are estimated using the reduction-based approach**. Various other issues pertaining to the general multi-level multidimensional rating estimation problem are discussed in the Appendix

1. **An overview of the Reduction-Based approach**

... (As we all know)

More important note:

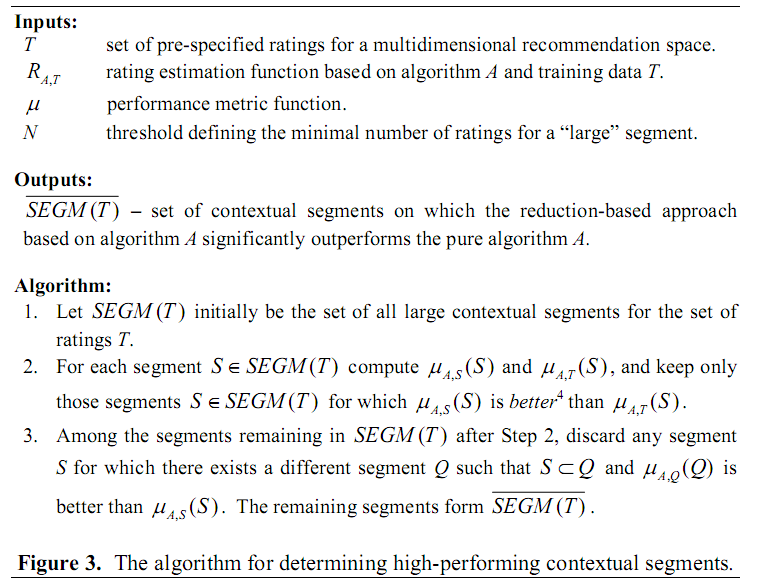
It is possible that it is better to use the reduction-based approach to recommend movies to see in the movie theaters on weekends and the traditional CF approach for movies to see at home on VCRs. This is the case because the reduction-based approach, on the one hand, focuses recommendations on a particular segment and builds a local prediction model for this segment, but, on the other hand, computes these recommendations based on a smaller number of points limited to the considered segment. This tradeoff between having more relevant data for calculating an unknown rating based only on the ratings with the same or similar context and having fewer data points used in this calculation belonging to a particular segment (i.e., the sparsity effect) explains why the reduction-based CF method can outperform traditional CF on some segments and underperform on others. Which of these two trends dominates on a particular segment may depend on the application domain and on the specifics of the available data. One solution to this problem is to combine the reduction-based and the traditional CF

approaches as explained in the next section.

* **Tóm ý cái này là trong 1 số trường hợp thì Reduction-Based vượt trội hơn Traditional 2D approach nhưng trong 1 số trường hợp ngược lại. Bởi vì có khả năng do giảm chiều sẽ dẫn đến Segment cuối cùng có quá ít dữ liệu có sẵn (Tưởng tượng có duy nhất 1 người đã rating trong hoàn cảnh hoàn toàn giống với user đang cần dự đoán), trong trường hợp này chấp nhận tính trên các Segment cấp trên bự hơn (nhiều case ko giống hết với ngữ cảnh hiện tại) nhưng như vậy sẽ lại có ý nghĩa hơn! Giải pháp là kết hợp Reduction-based và Traditional 2D.**

1. **Combined Reduction-Based and Traditional CF Approaches**

* Để so sánh Segment nào tốt hơn Segment nào cần có Performance Metric: MAE, MSE, F-Measure.
* Note that algorithm A can be an arbitrary two-dimensional rating estimation method, including collaborative filtering or any other heuristic-based and model-based methods discussed in Section 2. However, to illustrate how the reduction-based approach works, we will assume that A is a traditional collaborative filtering method in the remainder of this section and in our case study in Section 5
* **Chẳng cần sử dụng kĩ thuật nâng cấp mở rộng CF gì gì ở đây hết, chỉ cần 2D CF hết sức bình thường thôi. Lúc trước đọc ko kĩ cứ tưởng là Segment 2D thì mới dùng 2D CF được. Đọc kĩ lại ý tui tóm ở phần 2 về lý do phải kết hợp 2 kĩ thuật ta sẽ hiểu ngay vấn đề là cho dù Segment có mấy chiều hay thậm chí ko có Segment nào phù hợp phải tính trên toàn bộ Cube thì trong trường hợp này ta chỉ cần 2 chiều User và Item thôi, ko cần quan tâm ngữ cảnh làm j. Got it? Ví dụ Từ 5 chiều chọn được Segment tối ưu có 3 chiều, điều này có nghĩa là giảm được 2 chiều quan trọng đảm bảo việc dự đoán chính xác hơn nhưng nếu giảm thêm lần nữa thì sẽ ko có hiệu quả bằng nên dừng lại ở đây -> KO quan tâm tới mấy chiều (context) này nữa, tính trên User x Item luôn!**
* 2 pha của thuật toán:
* Pha 1: Chạy offline



* Pha 2: Chạy online

