Explainable Recommendation via Interpretable Feature Mapping

ABSTRACT

Trade-off exists between explainability and performance of the recommendations. To overcome this issue, we present a novel feature mapping approach that maps the uninterpretable latent features onto the interpretable aspect features, achieving both satisfactory performance and explainability in the recommendations by simultaneous minimization of recommendation and user preference losses.

1 INTRODUCTION

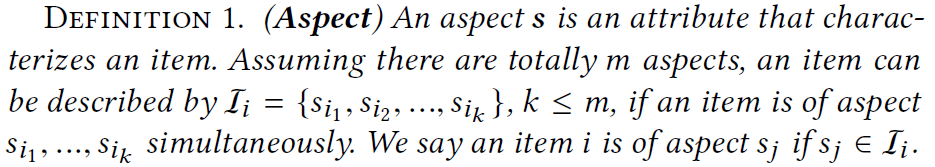
We implement this mapping strategy by designing an attentive multi-task learning approach.

We formulate the problem as: 1) how to find the interpretable aspect basis; 2) how to perform the interpretable feature mapping,

The main contributions of our work are: 1) We propose a novel feature mapping approach to map the complex uninterpretable features to interpretable aspect features, enabling explainable recommendations; 2) Borrowing strength across aspects, our model is capable of breaking the trade-off between recommendation performance and explainability with minimum auxiliary information; 3) Real-time explainable recommendation achieved by real-time

feature mapping.

2 RELATEDWORK

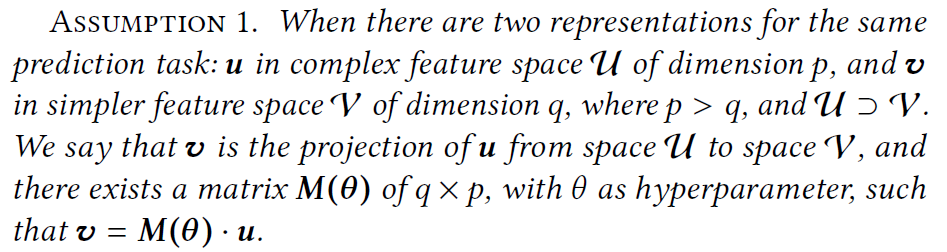


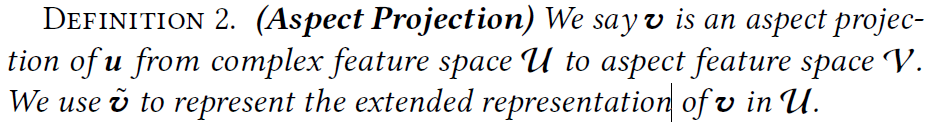
Latent Dirichlet Allocation (LDA)

Here we define the aspects directly using categorical information of an item and systematically optimize the recommendation performance and explainability using aspects. For example, we use movie genre (e.g., romance, thriller, suspension), or merchandise category (e.g., electronic, furniture, diary) as aspects.

3 THE PROPOSED MODEL

3.1 Motivations

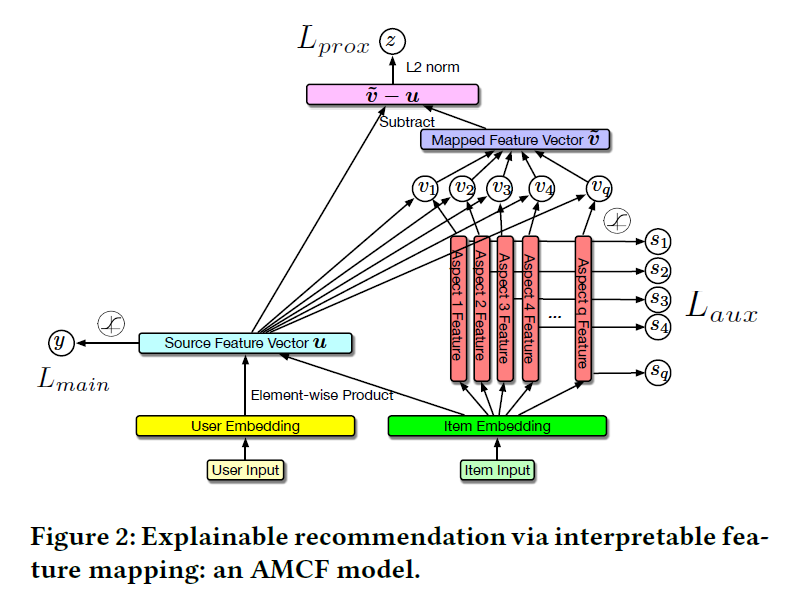




3.2 Aspect Feature Extraction

We select q predefined aspects as our interpretable features. From Assumption 1, to make the representation transformation from a source feature space U to an aspect feature space V, we need to first define the aspect space.

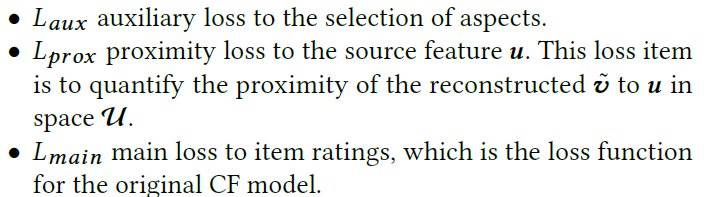
These aspect’s latent vectors can be inferred from auxiliary information such as user reviews, item profiles and tags. The aspects of an item is represented by multi-hot encoding.

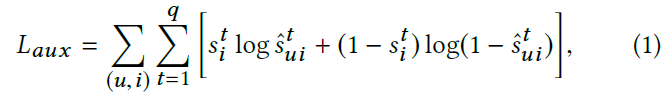


3.3 Aspect Projection

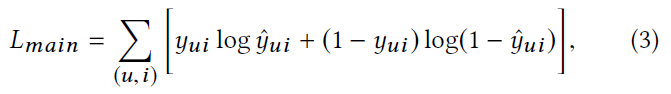
The use of attention mechanism is to create a feature mapping, i.e., M, from source space U to aspect space V.

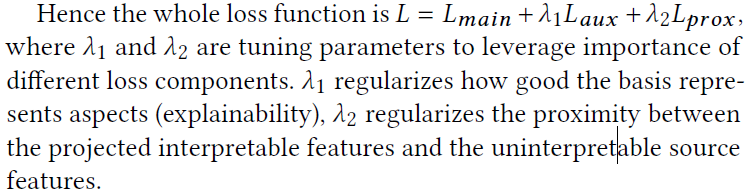
3.4 The Loss Function











5 CONCLUSION

In this paper, we proposed a novel interpretable feature mapping strategy attempting to achieve the two-folded goal: systems interpretability and recommendation explanation, without

the need for content-based auxiliary information.