Understanding Latent Factors Using a GWAP

Johannes Kunkel, Benedikt Loepp, Jürgen Ziegler

University of Duisburg-Essen, Duisburg, Germany
{firstname.lastname}@uni-due.de

ABSTRACT

Recommender systems relying on latent factor models often appear as black boxes to their users. Semantic descriptions for the factors might help to mitigate this problem. Achieving this automatically is, however, a non-straightforward task due to the models’ statistical nature. We present an output-agreement game that represents factors by means of sample items and motivates players to create such descriptions. A user study shows that the collected output actually reflects real-world characteristics of the factors.

KEYWORDS

Recommender Systems; Matrix Factorization; Game with a Purpose

ACM Reference Format:

1 INTRODUCTION AND RELATED WORK

Recommender Systems (RS) make it often difficult for users to understand the results, in particular when using model-based techniques such as Matrix Factorization (MF) [3]: While latent factor models are known for accuracy and efficiency, they are typically considered non-transparent [9]. Some works indicate that learned factors are related to actual real-world characteristics [3, 6], but only few steps have been taken to automatically explain the abstract dimensions, e.g. by associating them with mined topics [9] or tags [6]. Others visualized relations between predefined tags and factors [8] or reduced the dimensionality to display a map [4]. Still, making factor meanings explicit can be considered difficult, especially without predefined data or complex visualizations. Thus, it seems promising to count on voluntary user contribution. Games with a Purpose (GWAP), with their prominent method of Output-Agreement (OA) [11], are well known for motivating users to solve such a human computation problem. In OA, randomly matched pairs of players are presented with a common input and have to come up with the same output without any means of communication [11]. The winning strategy is thus to type in terms that describe the shared content as best as possible. Such games are often used to annotate images [11], but also for eliciting preferences in RS [2].

In this paper, we present a GWAP that follows the OA method for collecting semantic descriptions of latent factors.

2 THE GAME

In our game⁷, representative items are shown as shared input for one factor after the other. For producing the same output, players have to type in commonalities of these representatives. As descriptions, we collect all terms typed in for each factor. Such term-factor relations may then enable to e.g. explain MF results by showing terms related to factors relevant for the current recommendation.

Factor Model and Representatives: First, we use a Mahout ParallelSGDFactorizer for offline model training. With MovieLens 20M dataset, 10 factors, \( \lambda = 0.001 \) and 16 iterations, results were up to standard (\( \text{RMSE} = 0.86, \ NDCG@10 = 0.82 \)). While our approach is in principle independent of MF algorithm and parametrization, we deliberately choose a comparatively small number of factors. This is in line with earlier suggestions [7] and allows us to a) collect more terms per factor with less effort, and b) decrease likelihood of factors being redundant, thus diversifying game experience.

Next, for selecting sets of representatives both distinguishable and reflecting the semantics of model dimensions, we basically follow [7]: We calculate for each factor \( f \) and each movie \( i \) that has a value in the upper 25 % of values for \( f \) a score \( s_{if} = 0.4 \cdot \text{pop}(i) + 0.3 \cdot \text{rel}(i, f) + 0.3 \cdot \text{spec}(i, f) \), taking popularity (high number of ratings), relevance (high value for \( f \)), and specificity (high value for \( f \) but neutral for others) into account. For each factor, we then select the 25 movies with highest \( s_{if} \) as representatives. Weights and numbers are the result of pretesting.

Figure 1: LuckyGuy guesses (A) how his partner would describe the factor representatives (B). He made two guesses already (C). A leaderboard shows overall performances (D).

Game Mechanics: We implemented the game as a web application (Fig. 1), as recommended for OA [11]. A game lasts 3 min., during which two randomly matched players seek to play as many rounds as possible with the goal of gaining points and scoring high in a leaderboard. Rounds are randomly related to factors of the underlying MF model. In each round, 3 movies are randomly chosen from the 25 factor representatives, and displayed by means of poster, plot description, cast and director. A round ends a) as soon as a match is found in the terms entered by both players, i.e. they guess the same, or b) if both decide to skip e.g. because the movies are too hard to describe (leading to penalty points). Either way, they then proceed to the next round, i.e. an item set for another factor is shown.

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

LBRSLateRecSys ’18, October 2-7, 2018, Vancouver, BC, Canada
© 2018 Copyright held by the owner/author(s).

⁷Introduced in [5] in German language, see also: http://interactivesystems.info/lfg/.
3 EVALUATION

We conducted a user study to evaluate subjective game experience. We used 5-point Likert scales. We logged all interaction data, and observed how players entered guesses. Overall, similarities were very low ($\text{M} = 2.74$, $\text{SD} = 0.96$). This could be due to insufficient data, making items less distinguishable. On the other hand, factors might actually express similar aspects.

Some factors appear to have more obvious semantics: For instance, guess-match ratios in Table 1 show that players arrived at matches more often for factors 6 and 8, i.e., they are easier to describe.

Current game mechanics seem to favor rather general terms, e.g., genres such as "action" or "comedy." This effect has also been shown earlier [11], and can be prevented e.g., by taboo lists. However, implementing such mechanics would have required output data which we had not had prior to our study. On the other hand, very specific terms led to matches as well, e.g., "dog" for factor 2. This does not appear to result from participants describing a factor’s general meaning, but from certain movie posters being displayed. Yet, as the descriptions are only affected to a small degree, this issue will most likely vanish with more output data and X > 2.

4 CONCLUSIONS AND SHOWOUT

Study results show that our GWAP is fun to play and thus motivates users to produce output useful to better understand the hidden semantics of common RS models. Terms entered by players allow deriving meaningful and distinguishable latent factor descriptions, which may be used e.g., to explain recommendations by presenting keywords related to factors relevant for the active user and the respective items. Yet, investigating application areas is subject of future work, as is implementing a single player version as well as advanced game mechanics such as taboo lists for collecting further output data and more specific terms via gameplay.

REFERENCES